

A SIMPLE HASH CLASS WITH STRONG RANDOMNESS PROPERTIES IN GRAPHS AND HYPERGRAPHS*

MARTIN AUMÜLLER[†], MARTIN DIETZFELBINGER[‡], AND PHILIPP WOELFEL[§]

Abstract. We study randomness properties of graphs and hypergraphs generated by simple hash functions. Several hashing applications can be analyzed by studying the structure of d -uniform random (d -partite) hypergraphs obtained from a set S of n keys and d randomly chosen hash functions h_1, \dots, h_d by associating each key $x \in S$ with a hyperedge $\{h_1(x), \dots, h_d(x)\}$. Often it is assumed that h_1, \dots, h_d exhibit a high degree of independence. We present a simple construction of a hash class whose hash functions have small constant evaluation time and can be stored in sublinear space. We devise general techniques to analyze the randomness properties of the graphs and hypergraphs generated by these hash functions, and we show that they can replace other, less efficient constructions in cuckoo hashing (with and without stash), the simulation of a uniform hash function, the construction of a perfect hash function, generalized cuckoo hashing and different load balancing scenarios.

Key words. Hashing, cuckoo hashing, randomized algorithms, random graphs, load balancing

AMS subject classifications. 68P05, 68R10, 68W20, 05C80

1. Introduction. We study randomness properties of graphs and hypergraphs generated by hash functions of a particular simple structure. Consider a set S of n keys chosen from a finite set U , and a sequence $\vec{h} = (h_1, \dots, h_d)$, $d \geq 2$, of random hash functions $h_i: U \rightarrow [m] = \{0, \dots, m-1\}$ for some positive integer m . Then S and \vec{h} naturally define a d -partite d -uniform hypergraph $G(S, \vec{h}) := (V, E)$ with $V = V_{m,d}$, where $V_{m,d}$ is the union of d disjoint copies of $[m]$ and $E = \{(h_1(x), \dots, h_d(x)) \mid x \in S\}$.

Properties of such (hyper-)graphs are essential in the analysis of a number of randomized algorithms from a variety of applications, such as balanced allocation [67, 70], shared memory and PRAM simulations [52, 45], perfect hashing [38, 51], and recent hashing based dictionaries [35, 48, 60, 33, 47].

Often such algorithms are analyzed under the idealized *uniform hashing assumption*, which says that every hash function h employed by the algorithm is a truly random function. For example, if all functions h_1, \dots, h_d are truly random hash functions, then the graph $G(S, \vec{h})$ can be analyzed using the vast array of tools from random graph theory [5]. Since it is infeasible to store truly random functions, it is preferable to use small sets of hash functions (called *hash classes*), and sample random hash functions from those sets. These sets provide some weaker randomness guarantees. Carter and Wegman [9] defined a *universal hash class* \mathcal{H} as one which guarantees that any two distinct keys $x, x' \in U$ are mapped by a random function $h \in \mathcal{H}$ to the same function value only with a probability of $O(1/m)$. A stronger notion is that of a *k-wise independent* hash class \mathcal{H} [74], which says that for any k distinct keys x_1, \dots, x_k and

*This work was supported in part by DFG grant DI 412/10-1, DFG grant DI 412/10-2, and by a Discovery Grant from the National Sciences and Research Council of Canada (NSERC). Part of this work was done during a visit of the second author to MPI Saarbrücken, Germany, and during Dagstuhl seminars 07391, 08381, and 11121. This work encompasses the results of the conference papers [32], [77], and—to some extent—[2], and adds many new insights.

[†] IT University of Copenhagen, 2300 København, Denmark. (maau@itu.dk)

[‡] Fakultät für Informatik und Automatisierung, Technische Universität Ilmenau, 98694 Ilmenau, Germany. (martin.dietzfelbinger@tu-ilmenau.de)

[§] Department of Computer Science, University of Calgary, Calgary, Alberta T2N 1N4, Canada. (woelfel@cpsc.ucalgary.ca)

a random hash function $h \in \mathcal{H}$ the vector $(h(x_1), \dots, h(x_k))$ is uniformly distributed over $[m]^k$. The canonical representation of a k -independent hash class is the class of all degree $k - 1$ polynomials over some prime field. For the representation of such a polynomial, we just store its k coefficients (k words). The evaluation is possible in time $O(k)$.

Sometimes, ad-hoc analyses show that such limited randomness properties are sufficient for algorithms to exhibit the desired behavior. For example, Pagh, Pagh, and Ružić [59] showed that closed hashing with linear probing works well using only 5-wise independence. On the other hand, insufficient randomness can have subtle and unexpected negative effects on some algorithms. For example, Pătraşcu and Thorup [63] showed that linear probing behaves badly when used with a certain artificial class of 4-wise independent hash functions. In the same vein, it was experimentally observed in [60] and formally proved by Dietzfelbinger and Schellbach in [29, 30] that cuckoo hashing using the multiplicative class of hash functions from [24], although universal, does not work with high probability. For many other applications, such as cuckoo hashing [60] and ε -minwise independent hashing [42], we know that a logarithmic degree of independence suffices (in the size of the key set for the former, in $1/\varepsilon$ for the latter). In that case, polynomials use logarithmic space and evaluation time. If one aims for constant evaluation time, there is the construction by Siegel [69]—although Siegel states that his construction has constant albeit impractical evaluation time—and, more recently, the more efficient constructions by Thorup [71] and Christiani, Pagh, and Thorup [12].

Several techniques to circumvent the uniform hashing assumption have been proposed. The most general one is to “simulate” uniform hashing. The idea is to generate a class \mathcal{H} of hash functions at random such that for arbitrary given $S \subseteq U$ with high probability \mathcal{H} is “uniform” on S , which means that a random hash function $h \in \mathcal{H}$ restricted to the domain S is a truly random function. Such a simulation was presented by Pagh and Pagh in [56, 58], Dietzfelbinger and Rink [28], and in the precursor work [32]. However, such simulations require at least a linear (in $|S| \cdot \log m$) number of bits of additional space, which is often undesirable.

An alternative is the so-called *split-and-share* technique [35, 21, 28], in which S is first partitioned by a top-level hash function into smaller sets of keys, called bins. Then, a problem solution is computed for each bin, but all bins share the same hash functions. Since the size of each bin is significantly smaller than the size of S , it is possible to use a hash function that behaves like a truly random function on each bin. Finally, the problem solution of all bins is combined to a solution of the original problem. This technique cannot be applied uniformly to all applications, as ad-hoc algorithms depending on the application are required to merge the individual solutions for each bin to a solution of the original problem. In some scenarios, e.g., balanced allocation with high loads, the small deviations in the bin sizes incurred by the top-level hash function are undesirable. Moreover, additional costs in space and time are caused by the top-level splitting hash function and by compensating for a larger failure probability in each of the smaller bins.

Another perspective on uniform hashing is to assume that the key set $S = \{x_1, \dots, x_n\}$ itself is “sufficiently random”. Specifically, Mitzenmacher and Vadhan showed in [54] that when the distribution that governs $\{x_1, \dots, x_n\}$ has a low enough collision probability, then even using a hash function h from a 2-wise independent hash class \mathcal{H} makes the sequence $(h, h(x_1), \dots, h(x_n))$ distributed close to the uniform distribution on $\mathcal{H} \times R^n$ (see also [22]).

Besides these general techniques to circumvent the uniform hashing assumption, some research focuses on particular hash classes and their properties. Pătraşcu and Thorup [61] studied simple tabulation hashing, where each key is a tuple (x_1, \dots, x_c) which is mapped to the hash value¹ $(f_1(x_1) \oplus \dots \oplus f_c(x_c)) \bmod m$ by c uniform random hash functions f_1, \dots, f_c , each with a domain of cardinality $\lceil |U|^{1/c} \rceil$. The authors showed that simple tabulation hashing has striking randomness properties, and several applications (for example cuckoo hashing) exhibit good behavior if such hash functions are used. One year later, the same authors introduced “twisted tabulation hashing” [62], which gives even stronger randomness properties in many applications. Furthermore, Dahlgaard and Thorup proved that twisted tabulation is ε -minwise independent [17]. Very recently, Dahlgaard, Knudsen, Rotenberg, and Thorup extended the use of simple tabulation hashing to load balancing [16], showing that simple tabulation suffices for sequential load balancing with two choices. Simple tabulation hashing provides constant evaluation time with a description length that is polynomial (with exponent smaller than 1) in the size of the key set, just as with the hash functions studied in the present paper. Each application of tabulation hashing requires its own analysis. There are other approaches that trade higher evaluation time for smaller description length. For example, Reingold, Rothblum, and Wieder [65] showed that a class of hash functions introduced by Celis *et al.* [11] has strong enough randomness properties for running a slightly modified version of cuckoo hashing and sequential load balancing with two choices. While the hash class has non-constant evaluation time, its description length is notably smaller than what one gets using the standard polynomial approach for $\log n$ -wise independence ($O(\log n \log \log n)$ vs. $O(\log^2 n)$ bits) or tabulation hashing.

The Contribution. In this paper we focus on the properties of random graphs $G(S, \vec{h})$ generated by simple hash functions. These hash functions have been described before by Aumüller, Dietzfelbinger, and Woelfel in [3]. A function from their hash class, called \mathcal{Z} , combines simple k -independent hash functions with lookups in random tables. It can be evaluated efficiently in constant time, using a few arithmetic operations and table lookups. Each hash function can be stored in sublinear space, more precisely using $O(n^\gamma)$ bits for some $\gamma < 1$. To put our contribution in perspective, we first review some background. Building upon the work of Dietzfelbinger and Meyer auf der Heide [25], Aumüller, Dietzfelbinger, and Woelfel [3] showed that hash functions from class \mathcal{Z} have randomness properties strong enough to run cuckoo hashing with a stash with guarantees only known for fully random hash functions. To prove this result, Aumüller, Dietzfelbinger, and Woelfel studied the randomness properties of $G(S, h_1, h_2)$ when the hash function pair (h_1, h_2) is chosen randomly from \mathcal{Z} . They showed that the connected components of this graph behave, in some technical sense, very close to what is expected of the graph $G(S, h_1, h_2)$ when (h_1, h_2) is fully random.

Our contribution is that we provide a general framework that allows us to analyze applications whose analysis is based on arguments on the random graph described above when hash functions from \mathcal{Z} are used instead of fully random hash functions. To argue that the hash class can run a certain application, only random graph theory is applied, no details of the actual hash class need to be considered. Using this framework, we show that hash functions from \mathcal{Z} have randomness properties strong enough for many different applications, e.g., cuckoo hashing with a stash as described by Kirsch, Mitzenmacher, and Wieder in [49]; generalized cuckoo hashing as proposed by Fotakis, Pagh, Sanders, and Spirakis in [35] with two recently discovered insertion algorithms

¹ \oplus denotes the bit-wise XOR operation

due to Khosla [47] and Eppstein, Goodrich, Mitzenmacher and Pszozna [33] (in a sparse setting); the construction of a perfect hash function of Botelho, Pagh and Ziviani [7]; the simulation of a uniform hash function of Pagh and Pagh [58]; different types of load balancing as studied by Schickinger and Steger [67]. The analysis is done in a unified way which we hope will be of independent interest. We will find sufficient conditions under which it is possible to replace the full randomness assumption of a sequence of hash functions with explicit hash functions.

The General Idea. The analysis of hashing applications is often concerned with bounding (from above) the probability that random hash functions h_1, \dots, h_d map a given set $S \subseteq U$ of keys to some “bad” configuration of hash function values. These undesirable events can often be described by certain properties exhibited by the random graph $G(S, \vec{h})$. (Recall the notation $\vec{h} = (h_1, \dots, h_d)$.) For example, in cuckoo hashing a bad event occurs when $G(S, h_1, h_2)$ contains a very long simple path or a connected component with at least two cycles [60, 19].

If h_1, \dots, h_d are uniform hash functions, often a technique called *first moment method* (see, e.g., [5]) is employed to bound the probability of undesired events: In the standard analysis, one calculates the expectation of the random variable X that counts the number of subsets $T \subseteq S$ such that the subgraph $G(T, \vec{h})$ forms a “bad” substructure, as e.g., a connected component with two or more cycles. This is done by summing the probability that the subgraph $G(T, \vec{h})$ forms a “bad” substructure over all subsets $T \subseteq S$. One then shows that $E(X) = O(n^{-\alpha})$ for some $\alpha > 0$ and concludes that $\Pr(X > 0)$ —the probability that an undesired event happens—is at most $O(n^{-\alpha})$ by Markov’s inequality.

We state sufficient conditions allowing us to replace uniform hash functions h_1, \dots, h_d with hash function sequences from \mathcal{Z} without significantly changing the probability of the occurrence of certain undesired substructures $G(T, \vec{h})$. On a high level, the idea is as follows: We assume that for each $T \subseteq U$ we can split \mathcal{Z} into two disjoint parts: hash function sequences being *T-good*, and hash function sequences being *T-bad*. Choosing $\vec{h} = (h_1, \dots, h_d)$ at random from the set of *T-good* hash functions ensures that the hash values $h_i(x)$ with $x \in T$ and $1 \leq i \leq d$ are distributed fully randomly. Fix some set $S \subseteq U$. We identify some “exception set” $B_S \subseteq \mathcal{Z}$ (intended to be very small) such that for all $T \subseteq S$ we have: If $G(T, \vec{h})$ has an undesired property (e.g., a connected component with two or more cycles) and \vec{h} is *T-bad*, then $\vec{h} \in B_S$.

For $T \subseteq S$, disregarding the hash functions from B_S will allow us to calculate the probability that $G(T, \vec{h})$ has an undesired property as if \vec{h} were a sequence of fully random hash functions. Specifically, in [3, Lemma 2] it was already shown that

$$\Pr_{\vec{h} \in \mathcal{Z}}(X > 0) \leq E(X) + \Pr_{\vec{h} \in \mathcal{Z}}(B_S),$$

where the expectation is calculated assuming that \vec{h} is a pair of fully random hash functions. So, it is critical to find subsets B_S of sufficiently small probability. Whether or not this is possible depends on the substructures we are interested in. Here, we deviate from [3] and provide general criteria that allow us to bound the size of B_S from above entirely by using graph theory. This means that details about the hash function construction need not be known to argue that random hash functions from \mathcal{Z} can be used in place of uniform random hash functions for certain applications.

Outline and Suggestions. Section 2 introduces the considered class \mathcal{Z} of hash functions and provides the general framework of our analysis. Because of its abstract nature, the details of the framework might be hard to understand. A simple application

of the framework is provided in Section 2.4. There, we will discuss the use of hash class \mathcal{Z} in static cuckoo hashing. The reader might find it helpful to study the example first to get a feeling of how the framework is applied. Another way to approach the framework is to first read the paper [3]. This paper discusses one example of the framework with an application-specific focus, which might be easier to understand.

The following sections deal with applications of the hash function construction. Because these applications are quite diverse, the background of each one will be provided in the respective subsection right before the analysis.

Section 3 deals with randomness properties of \mathcal{Z} on (multi-)graphs. Here, Subsection 3.1 provides some groundwork for bounding the impact of using \mathcal{Z} in our applications. The subsequent subsections discuss the use of \mathcal{Z} in cuckoo hashing (with a stash), the simulation of a uniform hash function, the construction of a perfect hash function, and the behavior of \mathcal{Z} on connected components of $G(S, h_1, h_2)$.

The next section (Section 4) deals with applications whose analysis builds upon hypergraphs. As an introduction, we study generalized cuckoo hashing with $d \geq 3$ hash functions when the hash table load is low. Then, we consider two recently described alternative insertion algorithms for generalized cuckoo hashing. Finally, we prove that hash class \mathcal{Z} provides randomness properties strong enough for many different load balancing schemes.

In Section 5 we show how our analysis generalizes to the case that we use more involved hash functions as building blocks of hash class \mathcal{Z} , which lowers the total number of hash functions needed and the space consumption.

2. Basic Setup and Groundwork. Let U and R be two finite sets with $1 < |R| \leq |U|$. A *hash function with range R* is a mapping from U to R . In our applications, a hash function is applied on some key set $S \subseteq U$ with $|S| = n$. Furthermore, the range of the hash function is the set $[m] = \{0, \dots, m-1\}$ where often $m = \Theta(n)$. In measuring space, we always assume that $\log |U|$ is a term so small that it vanishes in big-Oh notation when compared with terms depending on n . If this is not the case, one first applies a hash function to collapse the universe to some size polynomial in n [69]. We say that a pair $x, y \in U, x \neq y$ *collides under a hash function g* if $g(x) = g(y)$.

The term *universal hashing*, introduced by Carter and Wegman in [9], refers to the technique of choosing a hash function at random from a *hash class* $\mathcal{H}_m \subseteq \{h \mid h: U \rightarrow [m]\}$.

DEFINITION 2.1 ([9, 10]). *For a constant $c \geq 1$, a hash class \mathcal{H} with functions from U to $[m]$ is called c -universal if for an arbitrary distinct pair of keys $x, y \in U$ we have*

$$\Pr_{h \in \mathcal{H}}(h(x) = h(y)) \leq c/m.$$

In our constructions we will use 2-universal classes of hash functions. Examples for c -universal hash classes can be found for example in [9, 24, 76]. In the following, \mathcal{F}_m^c denotes an arbitrary c -universal hash class with domain U and range $[m]$.

DEFINITION 2.2 ([74, 75]). *For an integer $\kappa \geq 2$, a hash class \mathcal{H} with functions from U to $[m]$ is called a κ -wise independent hash class if for arbitrary distinct keys $x_1, \dots, x_\kappa \in U$ and for arbitrary $j_1, \dots, j_\kappa \in [m]$ we have*

$$\Pr_{h \in \mathcal{H}}(h(x_1) = j_1 \wedge \dots \wedge h(x_\kappa) = j_\kappa) = 1/m^\kappa.$$

In other terms, choosing a hash function uniformly at random from a κ -wise independent class of hash functions guarantees that the hash values $h(x)$ are uniform in $[m]$ and

that the hash values of an arbitrary set of at most κ keys are independent. The classical construction of a κ -wise independent hash class is based on polynomials of degree $\kappa - 1$ over a finite field [74]. Another approach is to use tabulation-based hashing, see [72, 50, 61] for constructions using this approach. Tabulation-based constructions are often much faster in practice than polynomial-based hashing (cf. [72]) at the cost of using slightly more memory. Throughout this work, \mathcal{H}_m^κ denotes an arbitrary κ -wise independent hash class with domain U and range $[m]$.

We remark that Section 2.1 and Section 2.2 are quite natural generalizations of the basic definitions and observations made in [3] for pairs of hash functions. We give a full account for the convenience of the reader and to provide consistent notation.

2.1. The Hash Class. The hash class presented in this work draws ideas from many different papers. So, we first give a detailed overview of related work and key concepts.

Building upon the work on k -independent hash classes and two-level hashing strategies, e.g., the FKS-scheme of Fredman *et al.* [39], Dietzfelbinger and Meyer auf der Heide studied in [25, 26] randomness properties of hash functions from U to $[m]$ constructed in the following way: For given $k_1, k_2, m, n \geq 2$, and δ with $0 < \delta < 1$, set $\ell = n^\delta$. Let $f: U \rightarrow [m]$ be chosen from a k_1 -wise independent hash class, and let $g: U \rightarrow [\ell]$ be chosen from a k_2 -wise independent hash class. Fill a table $z[1..\ell]$ with random values from $[m]$. Given a key x , the hash function is evaluated as follows:

$$h(x) = f(x) + z[g(x)] \mod m.$$

For $m = n$, the hash class of [25] had many randomness properties that were only known to hold for fully random hash functions: When throwing n balls into n bins, where each candidate bin is chosen by “applying the hash function to the ball”, the expected maximum bin load is $O(\log n / \log \log n)$, and conditioned on a “good event” that occurs with probability $1 - \frac{1}{\text{poly}(n)}$ the probability that a bin contains $i \geq 1$ balls decreases exponentially with i . Other explicit hash classes that share this property were discovered by Pătraşcu and Thorup [61] and Celis *et al.* [11] only about two decades later.

Our work studies randomness properties of the same hash class as considered in Aumüller, Dietzfelbinger, and Woelfel [3]. For the convenience of the reader we define this class \mathcal{Z} next. It is a generalization of the hash class proposed in [25], modified so as to obtain pairs (h_1, h_2) of hash functions. One could choose two f -functions (from a k_1 -wise independent class), two z -tables, but *only one* g -function (from a k_2 -wise independent class) that is shared among h_1 and h_2 . In [3], this idea is further generalized so that for a given $c \geq 1$ one uses $2c$ z -tables and c g -functions.

We restrict the f -functions and the g -functions to be from very simple, 2-wise independent and 2-universal hash classes, respectively. This modification has two effects: it simplifies the analysis and it seems to yield faster hash functions in practice (see [3, Section 7]).

DEFINITION 2.3. *Let $c \geq 1$ and $d \geq 2$. For integers $m, \ell \geq 1$, and given $f_1, \dots, f_d: U \rightarrow [m]$, $g_1, \dots, g_c: U \rightarrow [\ell]$, and d two-dimensional tables $z^{(i)}[1..c, 0..\ell-1]$ with elements from $[m]$ for $i \in \{1, \dots, d\}$, we let $\vec{h} = (h_1, \dots, h_d) = (h_1, \dots, h_d)\langle f_1, \dots, f_d, g_1, \dots, g_c, z^{(1)}, \dots, z^{(d)} \rangle$, where*

$$h_i(x) = \left(f_i(x) + \sum_{1 \leq j \leq c} z^{(i)}[j, g_j(x)] \right) \mod m, \text{ for } x \in U, i \in \{1, \dots, d\}.$$

Let \mathcal{F}_ℓ^2 be an arbitrary two-universal class of hash functions from U to $[\ell]$, and let \mathcal{H}_m^2 be an arbitrary two-wise independent hash class from U to $[m]$. Then $\mathcal{Z}_{\ell,m}^{c,d}(\mathcal{F}_\ell^2, \mathcal{G}_m^2)$ is the class of all sequences $(h_1, \dots, h_d)(f_1, \dots, f_d, g_1, \dots, g_c, z^{(1)}, \dots, z^{(d)})$ for $f_i \in \mathcal{H}_m^2$ with $1 \leq i \leq d$ and $g_j \in \mathcal{F}_\ell^2$ with $1 \leq j \leq c$.

If arbitrary κ -wise independent hash classes are used as building blocks for the functions f_i , for $1 \leq i \leq d$, and g_j , for $1 \leq j \leq c$, one obtains the construction from [3]. However, the simpler hash functions are much easier to deal with in the proofs of this section. We defer the discussion of the general situation to Section 5.

While this is not reflected in the notation, we consider (h_1, \dots, h_d) as a structure from which the components g_1, \dots, g_c and $f_i, z^{(i)}$, $i \in \{1, \dots, d\}$, can be read off again. It is class $\mathcal{Z} = \mathcal{Z}_{\ell,m}^{c,d}(\mathcal{F}_\ell^2, \mathcal{G}_m^2)$ for some $c \geq 1$ and $d \geq 2$, made into a probability space by the uniform distribution, that we will study in the following. We usually assume that c and d are fixed and that m and ℓ are known. Also, the hash classes \mathcal{F}_ℓ^2 and \mathcal{G}_m^2 are arbitrary (if providing the necessary degree of universality or independence) and will not be mentioned explicitly below.

We will now discuss some randomness properties of hash class \mathcal{Z} . The central lemma is identical to [3, Lemma 1], but the proof is notably simpler because of the restriction to simpler hash functions as building blocks.

DEFINITION 2.4. For $T \subseteq U$, define the random variable d_T , the “deficiency” of $\vec{h} = (h_1, \dots, h_d)$ with respect to T , by $d_T(\vec{h}) = |T| - \max\{|g_1(T)|, \dots, |g_c(T)|\}$. Further, let

- (i) bad_T be the event that $d_T > 1$;
- (ii) good_T be $\overline{\text{bad}_T}$, i.e., the event that $d_T \leq 1$;
- (iii) crit_T be the event that $d_T = 1$.

Hash function sequences (h_1, \dots, h_d) in these events are called “ T -bad”, “ T -good”, and “ T -critical”, respectively.

It will turn out that if at least one of the functions g_j is injective on a set $T \subseteq U$, then all hash values on T are independent. The deficiency d_T of a sequence \vec{h} of hash functions measures how far away the hash function sequence is from this “ideal” situation. If \vec{h} is T -bad, then for each component g_j there are at least two “collisions” on T , i.e., there are at least two distinct pairs of keys from T that collide. If \vec{h} is T -good, then there exists a g_j -component with at most one collision on T . A hash function \vec{h} is T -critical if for all functions g_j there is at least one collision and there exists at least one function g_j such that g_j has exactly one collision on T . Note that the deficiency only depends on the g_j -components of a hash function. In the following, we will first fix these g_j -components when choosing a hash function. If $d_T(\vec{h}) \leq 1$ then the unfixed parts of the hash function, i.e., the entries in the tables $z^{(i)}$ and the f -functions, are sufficient to guarantee strong randomness properties of the hash function on T .

Our framework will build on the randomness properties of hash class \mathcal{Z} that are summarized in the next lemma. It comes in two parts. The first part makes the role of the deficiency of a hash function sequence from \mathcal{Z} precise, as described above. The second part states that for a fixed set $T \subseteq S$ three parameters govern the probability of the events crit_T or bad_T to occur: The size of T , the range $[\ell]$ of the g -functions, and their number. To be precise, this probability is at most $(|T|^2/\ell)^c$, which yields two consequences. When $|T|$ is much smaller than ℓ , the factor $1/\ell^c$ will make the probability of a hash function behaving badly on a small key set vanishingly small. But when $|T|^2$ is larger than ℓ , the influence of the failure term of the hash class is

significant. We will see later how to tackle this problem.

LEMMA 2.5. *Assume $d \geq 2$ and $c \geq 1$. For $T \subseteq U$ the following holds:*

- (a) *Conditioned on good_T (or on crit_T), the hash values $(h_1(x), \dots, h_d(x))$, $x \in T$, are distributed uniformly and independently in $[m]^d$.*
- (b) $\Pr(\text{bad}_T \cup \text{crit}_T) \leq (|T|^2/\ell)^c$.

Proof. Part (a): If $|T| \leq 2$, then h_1, \dots, h_d are fully random on T simply because f_1, \dots, f_d are drawn independently from a 2-wise independent hash class. So suppose $|T| > 2$. First, fix an arbitrary g -part of (h_1, \dots, h_d) so that crit_T occurs. (The statement follows analogously for good_T .) Let $j_0 \in \{1, \dots, c\}$ be such that there occurs exactly one collision of keys in T using g_{j_0} . Let $x, y \in T$, $x \neq y$, be this pair of keys (i.e., $g_{j_0}(x) = g_{j_0}(y)$). Arbitrarily fix all values in the tables $z^{(i)}[j, k]$ with $i \in \{1, \dots, d\}$, $j \neq j_0$, and $0 \leq k \leq \ell - 1$. Furthermore, fix $z^{(i)}[j_0, g_{j_0}(x)]$ with $i \in \{1, \dots, d\}$. The hash functions (h_1, \dots, h_d) are fully random on x and y since f_1, \dots, f_d are 2-wise independent. Furthermore, the function g_{j_0} is injective on $T - \{x, y\}$ and for each $x' \in (T - \{x, y\})$ the table cell $z^{(i)}[j_0, g_{j_0}(x')]$ is yet unfixed, for $i \in \{1, \dots, d\}$. Thus, the hash values $h_1(x'), \dots, h_d(x')$, $x' \in T - \{x, y\}$, are distributed fully randomly and are independent of the hash values of x and y .

Part (b): Assume $|T| \geq 2$. (Otherwise the events crit_T or bad_T cannot occur.) Suppose crit_T (or bad_T) is true. Then for each component g_i , $1 \leq i \leq c$, there are keys $x, y \in T$, $x \neq y$, such that $g_i(x) = g_i(y)$. Since g_i is chosen uniformly at random from a 2-universal hash class, the probability that such a pair exists is at most $\binom{|T|}{2} \cdot 2/\ell \leq |T|^2/\ell$. Since all g_i -components are chosen independently, the statement follows. \square

2.2. Graph Properties and the Hash Class. Here we describe how (hyper)graphs are built from a set of keys and a tuple of hash functions. The notation is superficially different from [3], the proof of the central lemma is identical.

We assume that the notion of a simple bipartite multigraph is known to the reader. A nice introduction to graph theory is given by Diestel [20]. We also consider *hypergraphs* (V, E) , which extend the notion of a graph by allowing edges to consist of more than two vertices, i.e., the elements of E are subsets of V of size 2 or larger. For an integer $d \geq 2$, a hypergraph is called *d-uniform* if each edge contains exactly d vertices. It is called *d-partite* if V can be split into d sets V_1, \dots, V_d such that no edge contains two vertices from the same class. A hypergraph (V', E') is a *subgraph* of a hypergraph (V, E) if $V' \subseteq V$ and if there is a one-to-one mapping $\varrho: E' \rightarrow E$ such that $e' \subseteq \varrho(e')$ for each $e' \in E'$. (The reader should be aware that while this definition gives the usual subgraphs for graphs, in the hypergraph setting it differs from the standard notion of a subhypergraph. To obtain a subgraph of a hypergraph in our sense, we may remove nodes and whole edges, but also delete nodes from single edges.) More notation for graphs and hypergraphs will be provided in Section 2.4 and Section 4, respectively.

We build graphs and hypergraphs from a set of keys $S = \{x_1, \dots, x_n\}$ and a sequence of hash functions $\vec{h} = (h_1, \dots, h_d)$, where $h_i: U \rightarrow [m]$ for $1 \leq i \leq d$, in the following way: The d -partite hypergraph $G(S, \vec{h}) = (V, E)$ has d copies of $[m]$ as vertex set and edge set $E = \{(h_1(x), \dots, h_d(x)) \mid x \in S\}$.² Also, the edge $(h_1(x_i), \dots, h_d(x_i))$

²In this paper, whenever we refer to a graph or a hypergraph we mean a multi-graph or multi-hypergraph, i.e., the edge set is a multiset. We also use the words “graph” and “hypergraph” synonymously in this section. Finally, note that our edges are tuples instead of sets to avoid problems with regard to the fact that the hash functions use the same range. The tuple notation (j_1, \dots, j_d) for edges is to be read as follows: j_1 is a vertex in the first copy of $[m]$, \dots , j_d is a vertex in the d -th

is labeled “ i ”.³ Since keys correspond to edges, the graph $G(S, \vec{h})$ has n edges and $d \cdot m$ vertices, which is the standard notation from a data structure point of view, but is non-standard in graph theory. For a set S and an edge-labeled graph G , we let $T(G) = \{x_i \mid x_i \in S, G \text{ contains an edge labeled } i\}$.

In the following, our main objective is to prove that with high probability certain subgraphs do not occur in $G(S, \vec{h})$. Formally, for $n, m, d \in \mathbb{N}$, $d \geq 2$, let $\mathcal{G}_{m,n}^d$ denote the set of all d -partite hypergraphs with vertex set $[m]$ in each class of the partition whose edges are labeled with distinct labels from $\{1, \dots, n\}$. A set $\mathbf{A} \subseteq \mathcal{G}_{m,n}^d$ is called a *graph property*. If for a graph G we have that $G \in \mathbf{A}$, we say that G *has property* \mathbf{A} . We shall always disregard isolated vertices.

For a key set S of size n , a sequence \vec{h} of hash functions from \mathcal{Z} , and a graph property $\mathbf{A} \subseteq \mathcal{G}_{m,n}^d$, we define the following random variables: For each $G \in \mathbf{A}$, let I_G be the indicator random variable that indicates whether G is a subgraph of $G(S, \vec{h})$ or not. (We demand the edge labels to coincide.) Furthermore, the random variable $N_S^{\mathbf{A}}$ counts the number of graphs $G \in \mathbf{A}$ which are subgraphs of $G(S, \vec{h})$, i.e., $N_S^{\mathbf{A}} = \sum_{G \in \mathbf{A}} I_G$.

Let \mathbf{A} be a graph property. Our main objective is then to estimate (from below) the probability that no subgraph of $G(S, \vec{h})$ has property \mathbf{A} . Formally, for given $S \subseteq U$ we wish to bound (from above)

$$\Pr_{\vec{h} \in \mathcal{Z}}(N_S^{\mathbf{A}} > 0). \quad (2.1)$$

In the analysis of a randomized algorithm, bounding (2.1) is often a classical application of the *first moment method*, which says that

$$\Pr_{\vec{h} \in \mathcal{Z}}(N_S^{\mathbf{A}} > 0) \leq \mathbb{E}_{\vec{h} \in \mathcal{Z}}(N_S^{\mathbf{A}}) = \sum_{G \in \mathbf{A}} \Pr_{\vec{h} \in \mathcal{Z}}(I_G = 1). \quad (2.2)$$

However, we cannot apply the first moment method directly to bound (2.1), since hash functions from \mathcal{Z} do not guarantee full independence on the key set, and thus the right-hand side of (2.2) is hard to calculate. However, we will prove an interesting connection to the expected number of subgraphs having property \mathbf{A} when the hash function sequence \vec{h} is fully random.

To achieve this, we will start by collecting “bad” sequences of hash functions. Intuitively, a sequence \vec{h} of hash functions is *bad* with respect to a key set S and a graph property \mathbf{A} if $G(S, \vec{h})$ has a subgraph G with $G \in \mathbf{A}$ and for the keys $T \subseteq S$ which form G the g -components of \vec{h} distribute T “badly”. (Recall the formal definition of “bad” from Definition 2.4.)

DEFINITION 2.6. For $S \subseteq U$ and a graph property \mathbf{A} let $B_S^{\mathbf{A}} \subseteq \mathcal{Z}$ be the event

$$\bigcup_{G \in \mathbf{A}} (\{I_G = 1\} \cap \text{bad}_{T(G)}).$$

This definition is slightly different from the corresponding definition in the paper [3, Definition 3], which considers one application of hash class \mathcal{Z} with an application-specific focus.⁴

³copy of $[m]$.

³We assume (w.l.o.g.) that the universe U is ordered and that each set $S \subseteq U$ of n keys is represented as $S = \{x_1, \dots, x_n\}$ with $x_1 < x_2 < \dots < x_n$.

⁴In [3] we defined $B_S^{\mathbf{A}} = \bigcup_{T \subseteq S} (\{G(T, \vec{h}) \text{ has property } \mathbf{A}\} \cap \text{bad}_T)$. This works well in the case

In addition to the probability space \mathcal{Z} together with the uniform distribution, we also consider the probability space in which we use d fully random hash functions from U to $[m]$, chosen independently. From here on, we will denote probabilities of events and expectations of random variables in the former case by \Pr and \mathbb{E} ; we will use \Pr^* and \mathbb{E}^* in the latter. The next lemma shows that for bounding $\Pr(N_S^A > 0)$ we can use $\mathbb{E}^*(N_S^A)$, i.e., the expected number of subgraphs having property A in the fully random case, and have to add the probability that the event B_S^A occurs. We call this additional summand the *failure term of \mathcal{Z} on A* .

LEMMA 2.7. *Let $S \subseteq U$ be given. For an arbitrary graph property A we have*

$$\Pr(N_S^A > 0) \leq \Pr(B_S^A) + \mathbb{E}^*(N_S^A). \quad (2.3)$$

Proof. We calculate:

$$\Pr(N_S^A > 0) \leq \Pr(B_S^A) + \Pr(\{N_S^A > 0\} \cap \overline{B_S^A}).$$

We only have to focus on the second term on the right-hand side. Using the union bound, we continue as follows:

$$\begin{aligned} \Pr(\{N_S^A > 0\} \cap \overline{B_S^A}) &\leq \sum_{G \in A} \Pr(\{I_G = 1\} \cap \overline{B_S^A}) \\ &= \sum_{G \in A} \Pr\left(\{I_G = 1\} \cap \left(\bigcap_{G' \in A} (\{I_{G'} = 0\} \cup \text{good}_{T(G')})\right)\right) \\ &\leq \sum_{G \in A} \Pr(\{I_G = 1\} \cap \text{good}_{T(G)}) \\ &\leq \sum_{G \in A} \Pr(I_G = 1 \mid \text{good}_{T(G)}) \\ &\stackrel{(i)}{=} \sum_{G \in A} \Pr^*(I_G = 1) = \mathbb{E}^*(N_S^A), \end{aligned}$$

where (i) holds by Lemma 2.5(b). \square

This lemma provides the strategy for bounding $\Pr(N_S^A > 0)$. The second summand in (2.3) can be calculated assuming full randomness and is often already known from the literature if the original analysis was conducted using the first moment method. The task of bounding the first summand is tackled separately in the next subsection.

2.3. A Framework for Bounding the Failure Term. As we have seen, using hash class \mathcal{Z} gives an additive failure term (cf. (2.3)) compared to the case that we bound $\Pr^*(N_S^A > 0)$ by the first moment method in the fully random case. Calculating $\Pr(B_S^A)$ looks difficult since we have to calculate the probability that there exists a subgraph G of $G(S, \vec{h})$ that has property A and where \vec{h} is $T(G)$ -bad. Since we know the probability that \vec{h} is $T(G)$ -bad from Lemma 2.5(b), we could tackle this task by

that we only consider randomness properties of the graph $G(S, h_1, h_2)$. In the hypergraph setting, “important” subgraphs of $G(S, \vec{h})$ often occur not in terms of the graph $G(T, \vec{h})$, for some set $T \subseteq S$, but by removing some vertices from the edges of $G(T, \vec{h})$. In Definition 2.6, we may consider exactly such subgraphs of $G(T, \vec{h})$ by defining A appropriately. The edge labels of a graph are used to identify which keys of S form the graph.

calculating the probability that there exists such a subgraph G under the condition that \vec{h} is $T(G)$ -bad, but then we cannot assume full randomness of \vec{h} on $T(G)$ to obtain a bound that a certain subgraph is realized by the hash values. We avoid this difficulty by taking another approach. We will find suitable events that contain B_S^A and where \vec{h} is guaranteed to behave well on the key set in question.

Observe the following relationship that is immediate from Definition 2.6.

LEMMA 2.8. *Let $S \subseteq U$, $|S| = n$, and let $A \subseteq B \subseteq \mathcal{G}_{m,n}^d$. Then $\Pr(B_S^A) \leq \Pr(B_S^B)$.*

We will now introduce two concepts that will allow us to bound the failure probability of \mathcal{Z} for “suitable” graph properties A .

DEFINITION 2.9 (Peelability). *A graph property A is called **peelable** if for all $G = (V, E) \in A$ with $|E| \geq 1$ there exists an edge $e \in E$ such that $(V, E - \{e\}) \in A$.*

An example for a peelable graph property for bipartite graphs, i.e., in the case $d = 2$, is the set of all connected bipartite graphs (disregarding isolated vertices), because removing an edge that lies on a cycle or an edge incident to a vertex of degree 1 does not destroy connectivity.

Peelable graph properties will help us in the following sense: Assume that B_S^A occurs, i.e., for the chosen $\vec{h} \in \mathcal{Z}$ there exists some graph $G \in A$ that is a subgraph of $G(S, \vec{h})$ and \vec{h} is $T(G)$ -bad. Let $T = T(G)$. In terms of the “deficiency” d_T of \vec{h} (cf. Definition 2.4) it holds that $d_T(\vec{h}) > 1$. If A is peelable, we can iteratively remove edges from G such that the resulting graphs still have property A . Let G' be a graph that results from G by removing a single edge. Then $d_{T(G)} - d_{T(G')} \in \{0, 1\}$. Eventually, because $d_\emptyset = 0$, we will obtain a subgraph $G' \in A$ of G such that \vec{h} is $T(G')$ -critical. In this case, we can again make use of Lemma 2.5(b) and bound the probability that G' is realized by the hash function sequence by assuming that the hash values are fully random.

However, peelability does not suffice to obtain low enough bounds for failure terms $\Pr(B_S^A)$; we need the following auxiliary concept, whose idea will become clear in the proof of the next lemma.

DEFINITION 2.10 (Reducibility). *Let $c \in \mathbb{N}$, and let A and B be graph properties. A is called **B-2c-reducible** if for all graphs $(V, E) \in A$ and sets $E^* \subseteq E$ with $|E^*| \leq 2c$ we have the following: There exists an edge set E' with $E^* \subseteq E' \subseteq E$ such that $(V, E') \in B$.*

If a graph property A is B-2c-reducible, we say that A *reduces to* B . The parameter c shows the connection to hash class \mathcal{Z} : it is the same parameter as the number of g_j -functions in hash class \mathcal{Z} .

To shorten notation, we let

$$\mu_t^A := \sum_{G \in A, |E(G)|=t} \Pr^*(I_G = 1)$$

be the expected number of subgraphs with exactly t edges having property A in the fully random case. The following lemma is the central result of this section and encapsulates our overall strategy to bound the additive failure term introduced by using hash class \mathcal{Z} instead of fully random hash functions.

LEMMA 2.11. *Let $c \geq 1$, $S \subseteq U$ with $|S| = n$, and let A , B , and C be graph properties such that $A \subseteq B$, B is a peelable graph property, and B reduces to C . Then*

$$\Pr(B_S^A) \leq \Pr(B_S^B) \leq \ell^{-c} \cdot \sum_{t=2}^n t^{2c} \cdot \mu_t^C.$$

Proof. By Lemma 2.8 we have $\Pr(B_S^A) \leq \Pr(B_S^B) = \Pr(\bigcup_{G \in \mathbf{B}} (\{I_G = 1\} \cap \text{bad}_{T(G)}))$. Assume that \vec{h} is such that B_S^B occurs. Then there exists a subgraph G of $G(S, \vec{h})$ such that $G \in \mathbf{B}$ and $d_{T(G)}(\vec{h}) > 1$. Fix such a graph.

Since \mathbf{B} is peelable, we iteratively remove edges from G until we obtain a graph $G' = (V, E')$ such that $G' \in \mathbf{B}$ and $\text{crit}_{T(G')}$ occurs. The latter is guaranteed, since $d_\emptyset(\vec{h}) = 0$ and since for two graphs G and G' , where G' results from G by removing a single edge, it holds that $d_{T(G)}(\vec{h}) - d_{T(G')}(\vec{h}) \in \{0, 1\}$. Since $\text{crit}_{T(G')}$ happens, for each g_i -component of \vec{h} , $1 \leq i \leq c$, there is at least one collision on $T(G')$. Furthermore, there exists one component g_{j_0} with $j_0 \in \{1, \dots, c\}$ such that exactly one collision on $T(G')$ occurs. For each g_i , $i \in \{1, \dots, c\}$, let x_i and y_i be two distinct keys such that G' contains the edges e_{x_i} and e_{y_i} labeled with these keys and such that x_i and y_i collide under g_i . Let $E^* = \bigcup_{1 \leq i \leq c} \{e_{x_i}, e_{y_i}\}$.

By construction $|E^*| \leq 2c$. Since \mathbf{B} reduces to \mathbf{C} , there exists some set E'' with $E^* \subseteq E'' \subseteq E'$ such that $G'' = (V, E'') \in \mathbf{C}$. By construction of E^* , each g_i -component has at least one collision on $T(G'')$. Moreover, g_{j_0} has exactly one collision on $T(G'')$. Thus, \vec{h} is $T(G'')$ -critical.

We calculate:

$$\begin{aligned} \Pr(B_S^A) &\leq \Pr(B_S^B) = \Pr\left(\bigcup_{G \in \mathbf{B}} (\{I_G = 1\} \cap \text{bad}_{T(G)})\right) \stackrel{(i)}{\leq} \Pr\left(\bigcup_{G' \in \mathbf{B}} (\{I_{G'} = 1\} \cap \text{crit}_{T(G')})\right) \\ &\stackrel{(ii)}{\leq} \Pr\left(\bigcup_{G'' \in \mathbf{C}} (\{I_{G''} = 1\} \cap \text{crit}_{T(G'')})\right) \leq \sum_{G'' \in \mathbf{C}} \Pr(\{I_{G''} = 1\} \cap \text{crit}_{T(G'')}) \\ &= \sum_{G'' \in \mathbf{C}} \Pr(I_{G''} = 1 \mid \text{crit}_{T(G'')}) \cdot \Pr(\text{crit}_{T(G'')}) \\ &\stackrel{(iii)}{\leq} \ell^{-c} \cdot \sum_{G'' \in \mathbf{C}} \Pr^*(I_{G''} = 1) \cdot |T(G'')|^{2c} \\ &= \ell^{-c} \cdot \sum_{t=2}^n \left(t^{2c} \cdot \sum_{\substack{G'' \in \mathbf{C} \\ |E(G'')|=t}} \Pr^*(I_{G''} = 1) \right) = \ell^{-c} \cdot \sum_{t=2}^n t^{2c} \cdot \mu_t^{\mathbf{C}}, \end{aligned}$$

where (i) holds because \mathbf{B} is peelable, (ii) is due to reducibility, and (iii) follows by Lemma 2.5. \square

We summarize the results of Lemma 2.7 and Lemma 2.11 in the following proposition.

PROPOSITION 2.12. *Let $c \geq 1$, $m \geq 1$, $S \subseteq U$ with $|S| = n$, and let \mathbf{A} , \mathbf{B} , and \mathbf{C} be graph properties such that $\mathbf{A} \subseteq \mathbf{B}$, \mathbf{B} is a peelable graph property, and \mathbf{B} reduces to \mathbf{C} . Assume that there are constants α, β such that*

$$\mathbf{E}^*(N_S^{\mathbf{A}}) := \sum_{t=1}^n \mu_t^{\mathbf{A}} = O(n^{-\alpha}), \quad (2.4)$$

and

$$\sum_{t=2}^n t^{2c} \mu_t^{\mathbf{C}} = O(n^\beta). \quad (2.5)$$

Then setting $\ell = n^{(\alpha+\beta)/c}$ and choosing \vec{h} at random from $\mathcal{Z}_{\ell,m}^{c,d}$ yields

$$\Pr(N_S^A > 0) = O(n^{-\alpha}).$$

Proof. The proposition follows immediately by plugging the failure probability bound from Lemma 2.11 into Lemma 2.7. \square

REMARK 2.13. *In the statement of Lemma 2.7 and Proposition 2.12 graph properties B and C can be the same graph properties, since every graph property reduces to itself.*

Proposition 2.12 shows the power of our framework. The conditions of this lemma can be checked without looking at the details of the hash functions, only by finding suitable graph properties that have a low enough expected number of subgraphs in the fully random case. Let us compare properties (2.4) and (2.5). Property (2.4) is the standard first moment method approach. So, it can often be checked from the literature whether a particular application seems suitable for an analysis with our framework or not. Property (2.5) seems very close to a first moment method approach, but there is one important difference to (2.4). The additional factor t^{2c} , coming from the randomness properties of the hash class, means that to obtain low enough bounds for (2.5), the average number of graphs with property C must decrease rapidly, e.g., exponentially, fast in t . This will be the case for almost all graph properties considered in this thesis.

In the analysis for application scenarios, we will use Lemma 2.7 and Lemma 2.11 instead of Proposition 2.12. Often, one auxiliary graph property suffices for many different applications and we think it is cleaner to first bound the failure term of \mathcal{Z} on this graph property using Lemma 2.11; then we only have to care about the fully random case and apply Lemma 2.7 at the end.

This concludes the development of the theoretical basis of this paper.

2.4. Step by Step Example: Analyzing Static Cuckoo Hashing. We start by fixing graph-related notation for usual graphs: We call an edge that is incident to a vertex of degree 1 a *leaf edge*. We call an edge a *cycle edge* if removing it does not disconnect any two nodes. A connected graph is called *acyclic* if it does not contain cycles. It is called *unicyclic* if it contains exactly one cycle.

Cuckoo hashing [60] is a well known dictionary algorithm that stores a (dynamically changing) set $S \subseteq U$ of size n in two hash tables, T_1 and T_2 , each of size $m \geq (1 + \varepsilon)n$ for some $\varepsilon > 0$. It employs two hash functions h_1 and h_2 with $h_1, h_2: U \rightarrow [m]$. A key x can be stored either in $T_1[h_1(x)]$ or in $T_2[h_2(x)]$, and all keys are stored in distinct table cells. Thus, to find or remove a key it suffices to check these two possible locations. For details on the insertion procedure we refer the reader to [60].

In this section, we deal with the static setting. Here the question is whether or not a key set S of size n can be stored in the two tables of size $(1 + \varepsilon)n$ each, for some $\varepsilon > 0$, using a pair of hash functions (h_1, h_2) according to the cuckoo hashing rules. To this end, we look at the bipartite graph $G(S, h_1, h_2)$ built from S and (h_1, h_2) . Recall that the vertices of G are two copies of $[m]$ and that each key $x_i \in S$ gives rise to an edge $(h_1(x), h_2(x))$ labeled i . If (h_1, h_2) allow storing S according to the cuckoo hashing rules, i.e., independent of the insertion algorithm, we call (h_1, h_2) *suitable* for S .

This section is meant as an introductory example for applying the framework. Already Pagh and Rodler showed in [60] that using a $\Theta(\log n)$ -wise independent hash

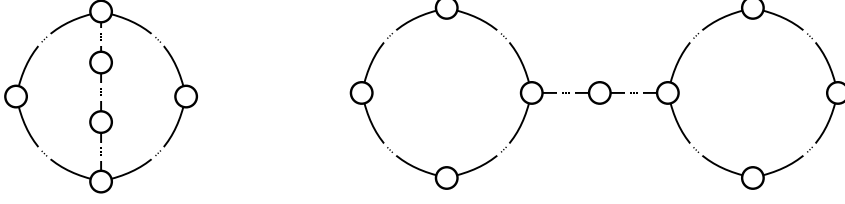


FIG. 2.1. The minimal obstruction graphs for cuckoo hashing, see also [19].

class suffices to run cuckoo hashing. Since standard cuckoo hashing is a special case of cuckoo hashing with a stash, the results here can also be proven using the techniques presented in [3]. However, the proofs here are notably simpler than the proofs needed for the analysis of cuckoo hashing with a stash. The central insight the reader should take away from the proof is that the analysis can be carried out by applying Lemma 2.11 alone, without reference to the inner structure of the hash functions.

We will prove the following theorem:

THEOREM 2.14. *Let $\varepsilon > 0$ and $0 < \delta < 1$ be given. Assume $c \geq 2/\delta$. For $n \geq 1$ consider $m \geq (1 + \varepsilon)n$ and $\ell = n^\delta$. Let $S \subseteq U$ with $|S| = n$. Then for (h_1, h_2) chosen at random from $\mathcal{Z} = \mathcal{Z}_{\ell, m}^{c, 2}$ the following holds:*

$$\Pr((h_1, h_2) \text{ is not suitable for } S) = O(1/n).$$

In the following, all statements of lemmas and claims use the parameter settings of Theorem 2.14.

It is not hard to see that (h_1, h_2) is suitable for S if and only if every connected component of $G(S, h_1, h_2)$ has at most one cycle [19]. So, if (h_1, h_2) is not suitable then $G(S, h_1, h_2)$ has a connected component with more than one cycle. This motivates considering the following graph property.

DEFINITION 2.15. *Let MOG (“minimal obstruction graphs”) be the set of all labeled graphs from $\mathcal{G}_{m, n}^2$ (disregarding isolated vertices) that form either a cycle with a chord or two cycles connected by a path of length $t \geq 0$.*

These two types of graphs form minimal connected graphs with more than one cycle, see Figure 2.1. So, if (h_1, h_2) is not suitable for S , then $G(S, h_1, h_2)$ contains a subgraph with property MOG. We summarize:

$$\Pr((h_1, h_2) \text{ is not suitable for } S) = \Pr(N_S^{\text{MOG}} > 0). \quad (2.6)$$

According to Lemma 2.7, the probability on the right-hand side of (2.6) is at most

$$\Pr(N_S^{\text{MOG}} > 0) \leq \Pr(B_S^{\text{MOG}}) + \mathbb{E}^*(N_S^{\text{MOG}}). \quad (2.7)$$

We first study the expected number of minimal obstruction graphs in the fully random case.

Bounding $\mathbb{E}^(N_S^{\text{MOG}})$.* The expected number of minimal obstruction graphs in the fully random case is well known from other work, see, e.g., [60]. The proof is included for the convenience of the reader and follows [60].

LEMMA 2.16.

$$\mathbb{E}^*(N_S^{\text{MOG}}) = O(1/m).$$

Proof. We start by counting unlabeled graphs with exactly t edges that form a minimal obstruction graph. Every minimal obstruction graph consists of a simple path of exactly $t - 2$ edges and two further edges which connect the endpoints of this path with vertices on the path. Since a minimal obstruction graph with t edges has exactly $t - 1$ vertices, there are no more than $(t - 1)^2$ unlabeled minimal obstruction graphs having exactly t edges. Fix an unlabeled minimal obstruction graph G . First, there are two ways to split the vertices of G into the two parts of the bipartition. When this is fixed, there are no more than m^{t-1} ways to label the vertices with labels from $[m]$, and there are no more than n^{t+1} ways to label the edges with labels from $\{1, \dots, n\}$. Fix such a fully labeled graph G' .

Now draw t labeled edges⁵ at random from $[m]^2$. The probability that these edges realize G' is exactly $1/m^{2t}$. We calculate:

$$\mathbb{E}^*(N_S^{\text{MOG}}) \leq \sum_{t=3}^n \frac{2n^t \cdot m^{t-1} \cdot (t-1)^2}{m^{2t}} \leq \frac{2}{m} \cdot \sum_{t=3}^n \frac{t^2 n^t}{m^t} = \frac{2}{m} \cdot \sum_{t=3}^n \frac{t^2}{(1+\varepsilon)^t} = O\left(\frac{1}{m}\right),$$

where the last step follows from the convergence of the series $\sum_{t=0}^{\infty} t^2/q^t$ for every $q > 1$. \square

Combining Lemma 2.16 and (2.7) we obtain:

$$\Pr(N_S^{\text{MOG}} > 0) \leq \Pr(B_S^{\text{MOG}}) + O\left(\frac{1}{m}\right). \quad (2.8)$$

It remains to bound the failure term $\Pr(B_S^{\text{MOG}})$.

Bounding $\Pr(B_S^{\text{MOG}})$. In the light of Definition 2.9, we first note that MOG is not peelable. So, we first find a peelable graph property that contains MOG. Since paths are peelable, and a minimal obstruction graph is “almost path-like” (*cf.* proof of Lemma 2.16), we relax the notion of a minimal obstruction graph in the following way.

DEFINITION 2.17. *Let RMOG (“relaxed minimal obstruction graphs”) consist of all graphs in $\mathcal{G}_{m,n}^2$ that form either (i) a minimal obstruction graph, (ii) a simple path, or (iii) a simple path and exactly one edge which connects an endpoint of the path with a vertex on the path. (We disregard isolated vertices.)*

By the definition, we obviously have that $\text{MOG} \subseteq \text{RMOG}$.

LEMMA 2.18. *RMOG is peelable.*

Proof. Let $G \in \text{RMOG}$. We may assume that G has at least two edges. We distinguish three cases:

Case 1: G is a minimal obstruction graph. Let G' be the graph that results from G when we remove an arbitrary cycle edge incident to a vertex of degree 3 or degree 4 in G . Then G' has property (iii) of Definition 2.17.

Case 2: G has property (iii) of Definition 2.17. Let G' be the graph that results from G when we remove an edge in the following way: If G contains a vertex of degree 3 then remove an arbitrary cycle edge incident to this vertex of degree 3, otherwise remove an arbitrary cycle edge. Then G' is a path and thus has property (ii) of Definition 2.17.

Case 3: G is a simple path. Let G' be the graph that results from G when we remove an endpoint of G with the incident edge. G' is a path and has property (ii) of Definition 2.17. \square

⁵The labels of these edges are equivalent to the edge labels of G' .

Standard cuckoo hashing is an example where we do not need every component of our framework, because there are “few enough” graphs having property RMOG to obtain low enough failure probabilities.

LEMMA 2.19.

$$\Pr(B_S^{\text{MOG}}) = O\left(\frac{n}{\ell^c}\right).$$

Proof. We aim to apply Lemma 2.11, where MOG takes the role of A and RMOG takes the role of B and C (cf. Remark 2.13), respectively, in the statement of that lemma.

CLAIM 2.20. *For $t \geq 2$, we have*

$$\mu_t^{\text{RMOG}} \leq \frac{6mt^2}{(1+\varepsilon)^t}.$$

Proof. We first count labeled graphs with exactly t edges having property RMOG. From the proof of Lemma 2.16 we know that there are fewer than $2 \cdot t^2 \cdot n^t \cdot m^{t-1}$ labeled graphs that form minimal obstruction graphs ((i) of Def. 2.17). Similarly, there are not more than $2 \cdot n^t \cdot m^{t+1}$ labeled paths ((ii) of Def. 2.17), and not more than $2 \cdot t \cdot n^t \cdot m^t$ graphs having property (iii) of Def. 2.17. Fix a labeled graph G with property RMOG having exactly t edges. Draw t labeled edges at random from $[m]^2$. The probability that these t edges realize G is exactly $1/m^{2t}$. We calculate:

$$\mu_t^{\text{RMOG}} \leq \frac{6t^2 n^t m^{t+1}}{m^{2t}} = \frac{6mt^2}{(1+\varepsilon)^t}. \quad \square$$

Using Lemma 2.11, we proceed as follows:

$$\Pr(B_S^{\text{MOG}}) \leq \ell^{-c} \cdot \sum_{t=2}^n t^{2c} \cdot \mu_t^{\text{RMOG}} \leq \ell^{-c} \cdot \sum_{t=2}^n \frac{6mt^{2(c+1)}}{(1+\varepsilon)^t} = O\left(\frac{n}{\ell^c}\right). \quad \square$$

Putting Everything Together. Plugging the results of Lemma 2.16 and Lemma 2.19 into (2.7) gives:

$$\Pr(N_S^{\text{MOG}} > 0) \leq \Pr(B_S^{\text{MOG}}) + \mathbb{E}^*(N_S^{\text{MOG}}) = O\left(\frac{n}{\ell^c}\right) + O\left(\frac{1}{m}\right).$$

Using that $m = (1+\varepsilon)n$ and setting $\ell = n^\delta$ and $c \geq 2/\delta$ yields Theorem 2.14.

This example gives insight into the situation in which our framework can be applied. The graph property under consideration (MOG) is such that the expected number of subgraphs with this property is polynomially small in n . The peeling process, however, yields graphs which are much more likely to occur, e.g., paths of a given length. The key in our analysis is finding suitable graph properties of “small enough” size. (That is the reason why the concept of “reducibility” from Definition 2.10 is needed in other applications: It makes the number of graphs that must be considered smaller.) The g -components of the hash functions from \mathcal{Z} provide a boost of ℓ^{-c} , which is then used to make the overall failure term again polynomially small in n .

The reader might find it instructive to apply Proposition 2.12 directly. Then, graph property MOG plays the role of graph property A in that proposition; graph property RMOG plays the role of B and C.

3. Applying the Framework to Graphs. In this section, we will study different applications of our hash class in algorithms and data structures whose analysis relies on properties of the graph $G(S, h_1, h_2)$. We shall study four different applications:

- A variant of cuckoo hashing called *cuckoo hashing with a stash* introduced by Kirsch, Mitzenmacher, and Wieder in [48].
- A construction for the simulation of a uniform hash function due to Pagh and Pagh [58].
- A construction of a (minimal) perfect hash function as described by Botelho, Pagh, and Ziviani [7].
- The randomness properties of hash class \mathcal{Z} on connected components of $G(S, h_1, h_2)$.

We start by studying the failure term of the hash class on a graph property that plays a central role in the applications.

3.1. Randomness Properties of Leafless Graphs. In this section we study the additive failure term of hash functions from \mathcal{Z} on a graph property that will be a key ingredient in the applications to follow. First, we recall some graph notation and present a counting argument from [3]. Subsequently, we study the failure term of \mathcal{Z} on the class of graphs which contain no leaf edges, so-called “leafless graphs”.

Leafless graphs are at the core of the analysis of many randomized algorithms and data structures, such as cuckoo hashing (note that the minimal obstruction graphs from Figure 2.1 have no leaves), the simulation of a uniform hash function as described by Pagh and Pagh in [58], and the construction of a perfect hash function from Botelho, Pagh, and Ziviani [7]. As we shall demonstrate in the subsequent sections, analyzing the case that we replace fully random hash functions by hash functions from \mathcal{Z} in these applications becomes easy when the behavior of the additive failure term of \mathcal{Z} on leafless graphs is known. The main result of this section says that the additive failure term of hash class $\mathcal{Z}_{\ell, m}^{c, 2}$ on leafless graphs is $O(n/\ell^c)$, and can thus be made as small as $O(n^{-\alpha})$ for $\ell = n^\delta$, $0 < \delta < 1$, and $c = \Theta(\alpha)$. On the way, we will use all steps of our framework developed in Section 2.

We recall some graph notation. The cyclomatic number $\gamma(G)$ is the dimension of the *cycle space* of a graph G . It is equal to the smallest number of edges we have to remove from G such that the remaining graph is a forest (an acyclic, possibly disconnected graph) [20]. Also, let $\zeta(G)$ denote the number of connected components of G (ignoring isolated vertices).

DEFINITION 3.1. *Let $N(t, \ell, \gamma, \zeta)$ be the number of unlabeled (multi-)graphs with ζ connected components and cyclomatic number γ that have $t - \ell$ inner edges and ℓ leaf edges.*

The following bound is central in our analysis; it is taken from [3, Lemma 4].

LEMMA 3.2. $N(t, \ell, \gamma, \zeta) = t^{O(\ell + \gamma + \zeta)}$.

We let $\text{LL} \subseteq \mathcal{G}_{m, n}^2$ consist of all bipartite graphs that contain no leaf edge. It will turn out that for all our applications LL will be a suitable “intermediate” graph property, i.e., for the graph property A interesting for the application it will hold $\text{A} \subseteq \text{LL}$, which will allow us to apply Lemma 2.8. (For example, graph property LL could have been used instead of graph property RMOG in the example of the previous section.) Hence our goal in this section is to show that there exists a constant $\alpha > 0$, which depends on the parameters ℓ and c of the hash class $\mathcal{Z}_{\ell, m}^{c, 2}$, such that

$$\Pr_{(h_1, h_2) \in \mathcal{Z}}(B_S^{\text{LL}}) = O(n^{-\alpha}).$$

Luckily, bounding $\Pr(B_S^{\text{LL}})$ is an example par excellence for applying Lemma 2.11. To use this lemma we have to find a suitable peelable graph property (note that LL is not peelable) and a suitable further graph property to which that graph property reduces.

We let LC consist of all graphs G from $\mathcal{G}_{m,n}^2$ that contain at most one connected component that has leaves, disregarding isolated vertices. If such a component exists, we call it the *leaf component* of G .

LEMMA 3.3. *LC is peelable.*

Proof. Suppose $G \in \text{LC}$ has at least one edge. If G has no leaf component then all edges are cycle edges, and removing an arbitrary edge leaves a cycle component or creates a leaf component. So, the resulting graph has property LC. If G has a leaf component C , remove a leaf edge. This makes the component smaller, but maintains property LC. So, the resulting graph has again property LC. \square

We will also need the following auxiliary graph property:

DEFINITION 3.4. *For $K \in \mathbb{N}$, let $\text{LCY}(K) \subseteq \mathcal{G}_{m,n}^2$ be the set of all bipartite graphs $G = (V, E)$ with the following properties (disregarding isolated vertices):*

1. *at most one connected component of G contains leaves (i.e., $\text{LCY}(K) \subseteq \text{LC}$);*
2. *the number $\zeta(G)$ of connected components is bounded by K ;*
3. *if present, the leaf component of G contains at most K leaf and cycle edges;*
4. *the cyclomatic number $\gamma(G)$ is bounded by K .*

LEMMA 3.5. *Let $c \geq 1$. Then LC is $\text{LCY}(4c)$ - $2c$ -reducible.*

Proof. Consider an arbitrary graph $G = (V, E) \in \text{LC}$ and an arbitrary edge set $E^* \subseteq E$ with $|E^*| \leq 2c$. We say that an edge that belongs to E^* is *marked*. G satisfies Property 1 of graphs from $\text{LCY}(4c)$. We process G in three stages:

Stage 1: Remove all components of G without marked edges. Afterwards at most $2c$ components are left, and G satisfies Property 2.

Stage 2: If G has a leaf component C , repeatedly remove unmarked leaf and cycle edges from C , while C has such edges. The remaining leaf and cycle edges in C are marked, and thus their number is at most $2c$; Property 3 is satisfied.

Stage 3: If there is a leaf component C with z marked edges (where $z \leq 2c$), then at least one of them is a leaf edge, and hence $\gamma(C) \leq z - 1$. Now consider an arbitrary leafless component C' with cyclomatic number z . We construct a suitable subgraph C'' of C' . For this, we need the following graph theoretic claim:

CLAIM 3.6. *Every leafless connected graph with i marked edges has a leafless connected subgraph with cyclomatic number $\leq i+1$ that contains all marked edges.*

Proof. Let $G = (V, E)$ be a leafless connected graph with i marked edges. If $\gamma(G) \leq i+1$, there is nothing to prove. So suppose $\gamma(G) \geq i+2$. Choose an arbitrary spanning tree (V, E_0) of G .

There are two types of edges in G : *bridge edges* and *cycle edges*. A bridge edge is an edge whose deletion disconnects the graph, cycle edges are those whose deletion does not disconnect the graph.

Clearly, all bridge edges are in E_0 . Let $E_{\text{mb}} \subseteq E_0$ denote the set of marked bridge edges. Removing the edges of E_{mb} from G splits V into $|E_{\text{mb}}| + 1$ connected components $V_1, \dots, V_{|E_{\text{mb}}|+1}$; removing the edges of E_{mb} from the spanning tree (V, E_0) will give exactly the same components. For each *cyclic* component V_j we choose one edge $e_j \notin E_0$ that connects two nodes in V_j . The set of these $|E_{\text{mb}}| + 1$ edges is called E_1 . Now each marked bridge edge lies on a path connecting two cycles in $(V, E_0 \cup E_1)$.

Recall from graph theory [20] the notion of a fundamental cycle: Clearly, each edge $e \in E - E_0$ closes a unique cycle with E_0 . The cycles thus obtained are called the fundamental cycles of G w. r. t. the spanning tree (V, E_0) . Each cycle in G can

be obtained as an XOR-combination of fundamental cycles. (This is just another formulation of the standard fact that the fundamental cycles form a basis of the “cycle space” of G , see [20].) From this it is immediate that every cycle edge of G lies on some fundamental cycle. Now we associate an edge $e' \notin E_0$ with each marked cycle edge $e \in E_{\text{mc}}$. Given e , let $e' \notin E_0$ be such that e is on the fundamental cycle of e' . Let E_2 be the set of all edges e' chosen in this way. Clearly, each $e \in E_{\text{mc}}$ is a cycle edge in $(V, E_0 \cup E_2)$.

Now let $G' = (V, E_0 \cup E_1 \cup E_2)$. Note that $|E_1 \cup E_2| \leq (|E_{\text{mb}}| + 1) + |E_{\text{mc}}| \leq i + 1$ and thus $\gamma(G') \leq i + 1$. In G' , each marked edge is on a cycle or on a path that connects two cycles. If we iteratively remove leaf edges from G' until no leaf is left, none of the marked edges will be affected. Thus, we obtain the desired leafless subgraph G^* with $\gamma(G^*) = \gamma(G') \leq i + 1$. \square

Claim 3.6 gives us a leafless subgraph C'' of our leafless component C' with $\gamma(C'') \leq z + 1$ that contains all marked edges of C' . We remove from G all vertices and edges of C' that are not in C'' . Doing this for all leafless components yields the final graph G . Summing contributions to the cyclomatic number of G over all (at most $2c$) components, we see that $\gamma(G) \leq 4c$; thus Property 4 is satisfied. \square

We now bound the additive failure term $\Pr(B_S^{\text{LL}})$.

LEMMA 3.7. *Let $S \subseteq U$ with $|S| = n$, $\varepsilon > 0$, $c \geq 1$, and let $\ell \geq 1$. Assume $m \geq (1 + \varepsilon)n$. If (h_1, h_2) are chosen at random from $\mathcal{Z}_{\ell, m}^{c, 2}$, then*

$$\Pr(B_S^{\text{LL}}) \leq \Pr(B_S^{\text{LC}}) = O(n/\ell^c).$$

Proof. According to Lemma 2.11 and Lemma 3.5 it holds that

$$\Pr(B_S^{\text{LL}}) \leq \Pr(B_S^{\text{LC}}) \leq \ell^{-c} \cdot \sum_{t=2}^n t^{2c} \cdot \mu_t^{\text{LCY}(4c)}.$$

CLAIM 3.8.

$$\mu_t^{\text{LCY}(4c)} = \frac{2n \cdot t^{O(1)}}{(1 + \varepsilon)^{t-1}}.$$

Proof. By Lemma 3.2, there are at most $t^{O(c)} = t^{O(1)}$ ways to choose a bipartite graph G in $\text{LCY}(4c)$ with t edges. Graph G cannot have more than $t + 1$ nodes, since cyclic components have at most as many nodes as edges, and in the single leaf component, if present, the number of nodes is at most one bigger than the number of edges. In each component of G , there are two ways to assign the vertices to the two sides of the bipartition. After such an assignment is fixed, there are at most m^{t+1} ways to label the vertices with elements of $[m]$, and there are not more than n^t ways to label the t edges of G with labels from $\{1, \dots, n\}$. Assume now such labels have been chosen for G . Draw t labeled edges according to the labeling of G from $[m]^2$ uniformly at random. The probability that they exactly fit the labeling of nodes and edges of G is $1/m^{2t}$. Thus,

$$\mu_t^{\text{LCY}(4c)} \leq \frac{2 \cdot m^{t+1} \cdot n^t \cdot t^{O(1)}}{m^{2t}} \leq \frac{2n \cdot t^{O(1)}}{(1 + \varepsilon)^{t-1}}. \quad \square$$

We use this claim to finish the proof of Lemma 3.7 by the following calculation:

$$\Pr(B_S^{\text{LC}}) \leq \ell^{-c} \sum_{t=2}^n t^{2c} \cdot \mu_t^{\text{LCY}(4c)} \leq \frac{2n}{\ell^c} \cdot \sum_{t=2}^n \frac{t^{O(1)}}{(1 + \varepsilon)^{t-1}} = O\left(\frac{n}{\ell^c}\right). \quad \square$$

We now turn our focus to cuckoo hashing with a stash. We reprove a result from [3] to demonstrate the power of the framework.

3.2. Cuckoo Hashing (with a Stash). Kirsch, Mitzenmacher, and Wieder [48] proposed augmenting the cuckoo hashing tables with a *stash*, an additional segment of storage that can hold up to s keys for some (constant) parameter s . They showed that using a stash of size s reduces the rehash probability to $\Theta(1/n^{s+1})$. For details of the algorithm, see [49].

We focus on the question whether the pair (h_1, h_2) allows storing the key set S in the two tables with a stash of size s . It is known from [49, 3] that a single parameter of $G = G(S, h_1, h_2)$ determines whether a stash of size s is sufficient to store S using (h_1, h_2) , namely the *excess* $\text{ex}(G)$. The excess $\text{ex}(G)$ of a graph G is defined as the minimum number of edges one has to remove from G so that all connected components of the remaining graph are acyclic or unicyclic. In [49] it is shown that the excess of a graph $G = (V, E)$ is $\text{ex}(G) = \gamma(G) - \zeta_{\text{cyc}}(G)$, where $\zeta_{\text{cyc}}(G)$ is the number of cyclic connected components in G . The connection between the excess of a graph and the failure probability of cuckoo hashing with a stash is that (h_1, h_2) are suitable for a key set S if and only if $\text{ex}(G(S, h_1, h_2)) \leq s$.

The following theorem shows that one can replace the full randomness assumption of [49] by hash functions from hash class \mathcal{Z} .

THEOREM 3.9 ([3]). *Let $\varepsilon > 0$ and $0 < \delta < 1$, let $s \geq 0$ be given. Assume $c \geq (s+2)/\delta$. For $n \geq 1$ consider $m \geq (1+\varepsilon)n$ and $\ell = n^\delta$. Let $S \subseteq U$ with $|S| = n$. Then for (h_1, h_2) chosen at random from $\mathcal{Z} = \mathcal{Z}_{\ell, m}^{c, 2}$ the following holds:*

$$\Pr(\text{ex}(G(S, h_1, h_2)) \geq s+1) = O(1/n^{s+1}).$$

Proof. As in [3], we define an *excess- $(s+1)$ core graph* as a leafless graph G with excess exactly $s+1$ in which all connected components have at least two cycles. By $\text{CG}(s+1)$ we denote the set of all excess- $(s+1)$ core graphs in $\mathcal{G}_{m, n}^2$. (For an illustration, see Figure 1 in [3].)

From [3, Lemma 6] we know that each $G = G(S, h_1, h_2)$ with $\text{ex}(G) \geq s+1$ contains an excess- $(s+1)$ core graph. Hence $\Pr(\text{ex}(G(S, h_1, h_2)) \geq s+1) \leq \Pr(N_S^{\text{CG}(s+1)} > 0)$. To prove Theorem 3.9, it suffices to show that $\Pr(N_S^{\text{CG}(s+1)} > 0) = O(1/n^{s+1})$. By Lemma 2.7, we know that

$$\Pr(N_S^{\text{CG}(s+1)} > 0) \leq \Pr(B_S^{\text{CG}(s+1)}) + \mathbb{E}^*(N_S^{\text{CG}(s+1)}). \quad (3.1)$$

From [3, Lemma 7] we know that $\mathbb{E}^*(N_S^{\text{CG}(s+1)}) = O(1/n^{s+1})$. Since $\text{CG}(s+1) \subseteq \text{LL}$, we may apply Lemma 3.7 and write

$$\Pr(N_S^{\text{CG}(s+1)} > 0) \leq O\left(\frac{n}{\ell^c}\right) + O\left(\frac{1}{n^{s+1}}\right) = O\left(\frac{1}{n^{s+1}}\right), \quad (3.2)$$

for the parameters used in Theorem 3.9. \square

3.3. Simulation of a Uniform Hash Function. Consider a universe U of keys and a finite set R . Suppose we want to construct a hash function that takes on fully random values from R on a key set $S \subseteq U$ of size n . The naïve construction just assigns a random hash value to each key $x \in S$ and stores the key-value pair in a

hash table that supports lookup in constant time and construction in expected time $O(n)$, e. g., cuckoo hashing (with a stash). For information theoretical reasons, this construction needs space at least $n \log |R|$. (See, e. g., [66, Lemma 5.3.1].) We will now see that we can achieve much more in (asymptotically) almost the same space.

By the term “*simulating uniform hashing for U and R* ” we mean an algorithm that does the following. On input $n \in \mathbb{N}$, a randomized procedure sets up a data structure DS_n that represents a hash function $h: U \rightarrow R$, which can then be evaluated efficiently for keys in U . For each set $S \subseteq U$ of cardinality n there is an event B_S that occurs with small probability such that conditioned on $\overline{B_S}$ the values $h(x)$, $x \in S$, are fully random. So, in contrast to the naïve construction from above, one h can be shared among many applications and works on each set $S \subseteq U$ of size n with high probability. The quality of the algorithm is determined by the space needed for DS_n , the evaluation time for h , and the probability of the event B_S , which we call the *failure probability of the construction*. It should be possible to evaluate h in constant time. Again, the information theoretical lower bound implies that at least $n \log |R|$ bits are needed to represent DS_n .

The first constructions that matched this space bound up to constant factors were proposed independently by Dietzfelbinger and Woelfel [32] and Östlin and Pagh [56]. In the following, let R be the range of the hash function to be constructed, and assume that (R, \oplus) is a commutative group. (For example, we could use $R = [t]$ with addition mod t .) We sketch the construction of [56] next.

The construction described in [56] builds upon the graph $G(S, h_1, h_2)$. Each vertex v of $G(S, h_1, h_2)$ is associated with a random element x_v from R . The construction uses a third hash function $h_3: U \rightarrow R$. All three hash functions have to be chosen from a n^δ -wise independent class. Let $x \in U$ be an arbitrary key and let (v, w) be the edge that corresponds to x in $G(S, h_1, h_2)$. The hash value of x is $h(x) = x_v \oplus x_w \oplus h_3(x)$. This construction uses $8n \cdot \log |R| + o(n) + O(\log \log |U|)$ bits of space and achieves a failure probability of $O(1/n^s)$ for each $s \geq 1$. (The influence of s on the description length of the data structure is in the $o(n) + O(\log \log |U|)$ term. It is also in the construction time of the hash functions h_1, h_2, h_3 .) The evaluation time is dominated by the evaluation time of the three highly-independent hash functions. The construction of [56] runs in time $O(n)$. In their full paper [58], a general method to reduce the description length of the data structure to $(1 + \varepsilon)n \log |R| + o(n) + O(\log \log |U|)$ bits was presented. This is essentially optimal. This technique adds a summand of $O(1/\varepsilon^2)$ to the evaluation time.

Another essentially space-optimal construction was presented by Dietzfelbinger and Rink in [28]. It is based on results of Calkin [8] and the “split-and-share” approach. It uses $(1 + \varepsilon)n \log |R| + o(n) + O(\log \log |U|)$ bits of space and has evaluation time $O(\max\{\log^2(1/\varepsilon), s^2\})$ for failure probability $O(n^{1-(s+2)/9})$.

The construction presented here is a modification of the construction in [58]. We replace the highly independent hash functions with functions from hash class \mathcal{Z} . The data structure consists of a hash function pair (h_1, h_2) from our hash class, two tables of size $m = (1 + \varepsilon)n$ each, filled with random elements from R , a two-wise independent hash function with range R , $O(s)$ small tables with entries from R , and $O(s)$ two-independent hash functions to pick elements from these tables. The evaluation time of h is $O(s)$, and for $S \subseteq U$, $|S| = n$, the event B_S occurs with probability $O(1/n^{s+1})$. The construction requires roughly twice as much space as the most space-efficient solutions [28, 58]. However, it seems to be a good compromise combining simplicity and fast evaluation time with moderate space consumption.

THEOREM 3.10. *Given $n \geq 1$, $0 < \delta < 1$, $\varepsilon > 0$, and $s \geq 0$, we can construct a data structure DS_n that allows us to compute a function $h: U \rightarrow R$ such that:*

- (i) *For each $S \subseteq U$ of size n there is an event B_S of probability $O(1/n^{s+1})$ such that conditioned on $\overline{B_S}$ the function h is distributed uniformly on S .*
- (ii) *For arbitrary $x \in U$, $h(x)$ can be evaluated in time $O(s/\delta)$.*
- (iii) *DS_n comprises $2(1 + \varepsilon)n \log |R| + o(n) + O(\log \log |U|)$ bits.*

Proof. Choose an arbitrary integer $c \geq (s + 2)/\delta$. Given U and n , set up DS_n as follows. Let $m = (1 + \varepsilon)n$ and $\ell = n^\delta$, and choose and store a hash function pair (h_1, h_2) from $\mathcal{Z} = \mathcal{Z}_{\ell, m}^{c, 2}$, with component functions g_1, \dots, g_c from \mathcal{F}_ℓ^2 . In addition, choose two random vectors $t_1, t_2 \in R^m$, c random vectors $y_1, \dots, y_c \in R^\ell$, and choose f at random from a 2-wise independent class of hash functions from U to R .

Using DS_n , the mapping $h: U \rightarrow R$ is defined as follows:

$$h(x) = t_1[h_1(x)] \oplus t_2[h_2(x)] \oplus f(x) \oplus y_1[g_1(x)] \oplus \dots \oplus y_c[g_c(x)].$$

DS_n satisfies (ii) and (iii) of Theorem 3.10. (If the universe is too large, it must be collapsed to size n^{s+3} first.) We show that it satisfies (i) as well. For this, let $S \subseteq U$ with $|S| = n$ be given.

First, consider only the hash functions (h_1, h_2) from \mathcal{Z} . By Lemma 3.7 we have $\Pr(B_S^{\text{LL}}) = O(n/\ell^c) = O(1/n^{s+1})$. Now fix $(h_1, h_2) \notin B_S^{\text{LL}}$, which includes fixing the components g_1, \dots, g_c . Let $T \subseteq S$ be such that $G(T, h_1, h_2)$ is the 2-core of $G(S, h_1, h_2)$, i.e., the maximal subgraph with minimum degree at least two. The graph $G(T, h_1, h_2)$ is leafless, and since $(h_1, h_2) \notin B_S^{\text{LL}}$, we have that (h_1, h_2) is T -good. Now we note that the part $f(x) \oplus \bigoplus_{1 \leq j \leq c} y_j[g_j(x)]$ of $h(x)$ acts exactly as one of our hash functions h_1 and h_2 , where f and y_1, \dots, y_c are yet unfixed. So, arguing as in the proof of Lemma 2.5 we see that h is fully random on T .

Now assume that f and the entries in the tables y_1, \dots, y_c are fixed. Following [58], we show that the random entries in t_1 and t_2 alone make sure that $h(x)$, $x \in S - T$, is fully random. For an idea of the proof let (x_1, \dots, x_p) be the keys in $S \setminus T$, ordered in such a way that the edge corresponding to x_i is a leaf edge in $G(T \cup \{x_1, \dots, x_i\}, h_1, h_2)$, for each $i \in \{1, \dots, p\}$. To obtain such an ordering, repeatedly remove leaf edges from $G = G(S, h_1, h_2)$, as long as this is possible. The sequence of corresponding keys removed in this way is x_p, \dots, x_1 . In [58] it is shown by an induction argument that h is uniform on $T \cup \{x_1, \dots, x_p\}$. \square

When this construction was first described in [2], it was the easiest to implement data structure to simulate a uniform hash function in almost optimal space. Nowadays, the construction of Pagh and Pagh can use the highly-independent hash class construction of Thorup [71] or Christiani, Pagh, and Thorup [12] instead of Siegel's construction. However, in the original analysis of Pagh and Pagh [58], the hash functions are required to be from an n^δ -wise independent hash class. It remains to be demonstrated by experiments that the construction of Pagh and Pagh in connection with the constructions mentioned above is efficient. We believe that using hash class \mathcal{Z} is much faster.

Applying the same trick as in [58], the data structure presented here can be extended to use only $(1 + \varepsilon)n$ words from R . The evaluation time of this construction is $O(\max\{\frac{1}{\varepsilon^2}, s\})$.

3.4. Construction of a (Minimal) Perfect Hash Function. A hash function $h: U \rightarrow [m]$ is perfect on $S \subseteq U$ if it is injective (or 1-on-1) on S . A perfect hash function is *minimal* if $|S| = m$. Here, S is assumed to be a static set. Perfect hash functions are usually applied when a large set of items is frequently queried; they allow

fast retrieval and efficient memory storage in this situation. We refer the reader to [7, 21] for surveys of constructions for (minimal) perfect hash functions.

The first explicit practical construction of a (minimal) perfect hash function which needs only $O(n)$ bits is due to Botelho, Pagh, and Ziviani [6] (full version [7]) and is based on a hypergraph approach. Our construction is based on this work. In [7], explicit hash functions based on the “split-and-share” approach were used. This technique builds upon a general strategy described by Dietzfelbinger in [21] and Dietzfelbinger and Rink in [28] to make the “full randomness assumption” feasible in the construction of a perfect hash function. Botelho *et al.* showed in experiments that their construction is practical, even when realistic hash functions are used. Our goal is to show that hash functions from class \mathcal{Z} can be used in a specific version of their construction as well. After proving the main result, we will speculate about differences in running time between the split-and-share approach of [7] and hash class \mathcal{Z} .

The construction of [7] to build a perfect hash function mapping keys from a key set S to $[2m]$ with $m = (1 + \varepsilon)n$ works as follows. First, the graph $G(S, h_1, h_2)$ is built. If this graph contains a cycle, new hash functions are chosen and the graph is built anew. If the graph is acyclic, a peeling algorithm is used to construct a one-to-one mapping $\sigma: S \rightarrow [2m]$ with $\sigma(x) \in \{h_1(x), m + h_2(x)\}$ for all $x \in S$. The result of the peeling procedure also makes it possible to construct two tables $t_1[0..m-1]$ and $t_2[0..m-1]$ storing bits, with the property that

$$\sigma(x) = \begin{cases} h_1(x) \\ m + h_2(x) \end{cases} \iff t_1[h_1(x)] \oplus t_2[h_2(x)] = \begin{cases} 0 \\ 1 \end{cases}, \text{ for } x \in S.$$

It is then obvious how $\sigma(x)$ can be calculated in constant time, given $t_1[0..m-1]$ and $t_2[0..m-1]$.

If (h_1, h_2) are fully random hash functions, the probability that the graph is acyclic, i.e., the probability that the construction succeeds, for $m = (1 + \varepsilon)n$ is

$$\sqrt{1 - \left(\frac{1}{1 + \varepsilon}\right)^2}. \quad (3.3)$$

We replace the pair (h_1, h_2) of hash functions by functions from \mathcal{Z} . The next lemma shows that for $m \geq 1.08n$ we can build a perfect hash function for a key set S by trying the construction of Botelho *et al.* a constant number of times (in expectation).

LEMMA 3.11. *Let $S \subseteq U$ with $|S| = n$. Let $\varepsilon \geq 0.08$, and let $m \geq (1 + \varepsilon)n$. Set $\ell = n^\delta$ and $c \geq 1.25/\delta$. Then for a randomly chosen pair $(h_1, h_2) \in \mathcal{Z}_{\ell, m}^{c, 2}$ we have*

$$\Pr(G(S, h_1, h_2) \text{ is acyclic}) \geq 1 + \frac{1}{2} \ln \left(1 - \left(\frac{1}{1 + \varepsilon} \right)^2 \right) - o(1). \quad (3.4)$$

Figure 3.1 depicts the difference between the probability bounds in (3.3) and in (3.4). The theoretical bound using a first moment approach is close to the behavior in a random graph when $\varepsilon \geq 1$.

Proof. [Of Lemma 3.11] Let CYC be the set of all cycles in $\mathcal{G}_{m, n}^2$. Note that all these cycles have even length, since we consider bipartite graphs. By Lemma 2.7, we may bound $\Pr(N_S^{\text{CYC}} > 0)$ by $\Pr(B_S^{\text{CYC}}) + E^*(N_S^{\text{CYC}})$. Since $\text{CYC} \subseteq \text{LL}$, we know that $\Pr(B_S^{\text{CYC}}) = O(n/\ell^c)$, see Lemma 3.7. For the parameter choices $\ell = n^\delta$ and $c \geq 1.25/\delta$ we have $\Pr(B_S^{\text{CYC}}) = o(1)$. We now focus on the second summand and

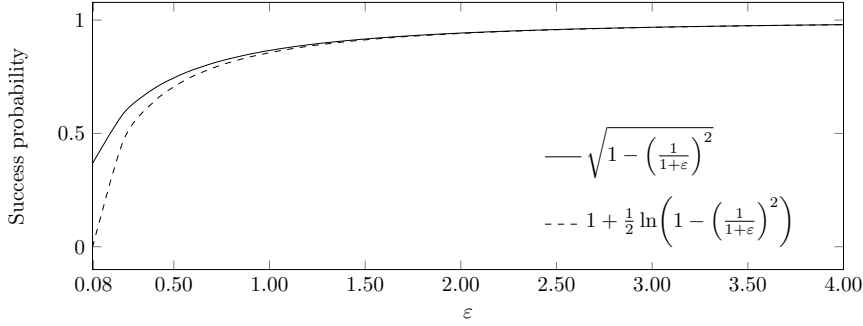


FIG. 3.1. Comparison of the probability of a random graph being acyclic and the theoretical bound following from a first moment approach for values $\varepsilon \in [0.08, 4]$.

calculate (as in [7]):

$$\begin{aligned} \mathbb{E}^*(N_S^{\text{CYC}}) &= \sum_{t=1}^{n/2} \mu_{2t}^{\text{CYC}} \leq \sum_{t=1}^{n/2} \frac{\binom{n}{2t} (2t)! \cdot m^{2t}}{2t \cdot m^{2 \cdot 2t}} = \sum_{t=1}^{n/2} \frac{\binom{n}{2t} \cdot (2t)!}{2t \cdot m^{2t}} \leq \sum_{t=1}^{n/2} \frac{n^{2t}}{2t \cdot m^{2t}} \\ &= \sum_{t=1}^{n/2} \frac{1}{2t \cdot (1+\varepsilon)^{2t}} \leq \sum_{t=1}^{\infty} \frac{1}{2t \cdot (1+\varepsilon)^{2t}} = -\frac{1}{2} \ln \left(1 - \left(\frac{1}{1+\varepsilon} \right)^2 \right), \end{aligned}$$

where the last step is the Maclaurin expansion. \square

According to Lemma 3.11, we can build a perfect hash function with range $[2.16n]$ with a constant number of constructions of $G(S, h_1, h_2)$ (in expectation). To store the data structure we need $2.16n$ bits (to store \mathbf{g}_1 and \mathbf{g}_2), and $o(n)$ bits to store the pair (h_1, h_2) from \mathcal{Z} . For example, for a set of $n = 2^{32}$ keys, i.e., about 4.3 billion keys, the pair (h_1, h_2) may consist of ten tables with 256 entries each, five 2-universal hash functions, and two 2-independent hash functions, see Lemma 3.11 with parameters $c = 5$ and $\delta = 1/4$. This seems to be more practical than the split-and-share approach from [7] which uses more and larger tables per hash function, cf. [7, Section 4.2]. However, it remains future work to demonstrate in experiments how both approaches compare to each other. To obtain a minimal perfect hash function, one has to compress the perfect hash function further. This roughly doubles the description length, see [7] for details.

In their paper [7], Botelho *et al.* showed that minimal space usage is achieved when using three hash functions h_1, h_2, h_3 to build the hypergraph $G(S, h_1, h_2, h_3)$. In this case, one can construct a perfect hash function with range $[1.23n]$ with high probability. Since the g -values must then index three hash functions, $1.23n \cdot \log_2 3 \approx 1.95n$ bits are needed to store the bit vectors. According to [7], the minimal perfect hash function needs about $2.62n$ bits. Our results with regard to hypergraphs do not lead to a construction that can compete.

3.5. Connected Components of $G(S, h_1, h_2)$ are small. As is well known from the theory of random graphs, for a key set $S \subseteq U$ of size n and $m = (1 + \varepsilon)n$, for $\varepsilon > 0$, and fully random hash functions $h_1, h_2 : U \rightarrow [m]$ the graph $G(S, h_1, h_2)$ w.h.p. contains only components of at most logarithmic size which are trees or unicyclic. (This is the central argument for standard cuckoo hashing to work.) We show here that hash class \mathcal{Z} can provide this behavior if one is willing to accept a density that is smaller by a constant factor. Such situations have been considered in the seminal

work of Karp *et al.* [46] on the simulation of shared memory in distributed memory machines.

We give the following result as a corollary. It has first appeared in [46], for a different class of (\sqrt{n} -wise independent) hash functions. Here we prove it for hash class \mathcal{Z} .

COROLLARY 3.12 ([46, Lemma 6.3]). *Let $S \subseteq U$ with $|S| = n$. Let $m \geq 6n$. Then for each $\alpha \geq 1$, there are $\beta, \ell, c, s \geq 1$ such that for $G = G(S, h_1, h_2)$ with $(h_1, h_2) \in \mathcal{Z}_{\ell, m}^{c, 2}$ we have that*

- (a) $\Pr(G \text{ has a connected component with at least } \beta \log n \text{ vertices}) = O(n^{-\alpha})$.
- (b) $\Pr(G \text{ has a connected component with } k \text{ vertices and } \geq k + s - 1 \text{ edges}) = O(n^{-\alpha})$.

Proof. We start with the proof of (b). If $G = G(S, h_1, h_2)$ has a connected component A with k vertices and at least $k + s - 1$ edges, then $\text{ex}(G) \geq s - 1$. According to Theorem 3.9, the probability that such a component appears is $O(1/n^\alpha)$, for $s = \alpha$ and $c \geq 2(\alpha + 1)$.

The proof of Part (a) requires more care. For this part, we may focus on the probability that G contains a tree with $k = \beta \log n$ vertices. We let T consist of all trees with k vertices in $\mathcal{G}_{m, n}$ and apply Lemma 2.7 to get

$$\Pr(N_S^{\mathsf{T}} > 0) \leq \Pr(B_S^{\mathsf{T}}) + E^*(N_S^{\mathsf{T}}). \quad (3.5)$$

Since $\mathsf{T} \subseteq \text{LCY}$, we have that $\Pr(B_S^{\mathsf{T}}) = O(n/\ell^c)$, see Lemma 3.7. We now bound the second summand of (3.5). Note that the calculations are essentially the same as the ones made in [46] to prove their Lemma 6.3. By Cayley's formula we know that there are k^{k-2} labeled trees with vertex set $\{1, \dots, k\}$. Fix such a tree T^* . We can label the edges of T^* with $k - 1$ keys from S in $\binom{n}{k-1} \cdot (k-1)!$ many ways. Furthermore, there are two ways to fix which vertices of T^* belong to which side of the bipartition. After this, there are not more than $\binom{2m}{k}$ ways to assign the vertices of T^* to vertices in the bipartite graph $G(S, h_1, h_2)$. Once all these labels of T^* are fixed, the probability that the hash values of (h_1, h_2) realize T^* is $1/m^{2(k-1)}$. We can thus calculate:

$$\begin{aligned} E^*(N_S^{\mathsf{T}}) &\leq \frac{\binom{n}{k-1} \cdot k^{k-2} \cdot 2 \cdot (k-1)! \cdot \binom{2m}{k}}{m^{2(k-1)}} \leq \frac{2^{k+1} \cdot m^2 \cdot k^{k-2}}{6^k \cdot k!} \\ &\leq \frac{2^{k+1} \cdot m^2 \cdot k^{k-2}}{6^k \cdot \sqrt{2\pi k} \cdot (k/e)^k} \leq 2m^2 \cdot \left(\frac{e}{3}\right)^k \end{aligned}$$

Part (a) follows for $k = \Omega(\log n)$. \square

The authors of [46] use the hash functions from [25] (precursor of \mathcal{Z}) combined with functions from Siegel's hash class to obtain a hash class with high (\sqrt{n} -wise) independence. They need this high level of independence in the proof of their Lemma 6.3, which states properties of the connected components in the graph built from the key set and these highly independent hash functions. Replacing Lemma 6.3 in [46] with our Corollary 3.12 immediately implies that the results of [46], in particular, their Theorem 6.4, also hold when (only) hash functions from \mathcal{Z} are used. In particular, in [46] the sparse setting where m is at least $6n$ was considered as well. We remark that statement (b) of Corollary 3.12 holds for $m \geq (1 + \varepsilon)n$.

Moreover, this result could be applied to prove results for cuckoo hashing (with a stash), and de-amortized cuckoo hashing of Arbritman *et al.* [1]. However, note that while in the fully random case the statement of Corollary 3.12 holds for $m = (1 + \varepsilon)n$, here we had to assume $m \geq 6n$, which yields only very low hash table load. We

note that this result cannot be improved to $(1 + \varepsilon)n$ using the first moment approach inherent in our approach and the approach of [46] (for \sqrt{n} -wise independence), since the number of unlabeled trees that have to be considered in the first moment approach is too large [57]. It remains open to show that graphs built with our class of hash functions have small connected components for all $\varepsilon > 0$.

4. Applying the Framework on Hypergraphs. In this section we will discuss some applications of hash class \mathcal{Z} in the setting with more than two hash functions, i.e., each edge of $G(S, \vec{h})$ contains at least three vertices. We will study three different applications: Generalized cuckoo hashing with $d \geq 3$ hash functions as proposed by Fotakis, Pagh, Sanders, and Spirakis [35], two recently described insertion algorithms for generalized cuckoo hashing due to Khosla [47] and Eppstein, Goodrich, Mitzenmacher, and Psziona [33], and different schemes for load balancing as studied by Schickinger and Steger [67].

For applications regarding generalized cuckoo hashing, we will study the failure term of \mathcal{Z} on the respective graph properties directly. We will show that \mathcal{Z} allows running these applications efficiently. However, we have to assume that the load of the hash table is rather low. For the application with regard to load balancing schemes, the failure term of \mathcal{Z} will be analyzed by means of a very general graph property. However, it requires higher parameters when setting up a hash function from \mathcal{Z} , which degrades the performance of these hash functions. We will start by introducing some notation and making a small generalization to the framework.

Hypergraph Notation. A hypergraph extends the notion of an undirected graph by allowing edges to consist of more than two vertices. We use the hypergraph notation from [68, 44]. A hypergraph is called *d-uniform* if every edge contains exactly d vertices. Let $H = (V, E)$ be a hypergraph. A *hyperpath* from u to v in H is a sequence $(u = u_1, e_1, u_2, e_2, \dots, e_{t-1}, u_t = v)$ such that $e_i \in E$ and $u_i, u_{i+1} \in e_i$, for $1 \leq i \leq t-1$. The hypergraph H is *connected* if for each pair of vertices $u, v \in V$ there exists a hyperpath from u to v .

The *bipartite representation* of a hypergraph H is the bipartite graph $\text{bi}(H)$ where vertices of H are the vertices on the right side of the bipartition, the edges of H correspond to vertices on the left side of the bipartition, and two vertices are connected by an edge in the bipartite graph if the corresponding edge in the hypergraph contains the corresponding vertex. Note that a hypergraph H' is a subgraph of a hypergraph H (as defined in Section 2) if and only if $\text{bi}(H')$ is a subgraph of $\text{bi}(H)$ in the standard sense.

We will use a rather strict notion of cycles in hypergraphs. A connected hypergraph is called a *hypertree* if $\text{bi}(H)$ is a tree. A connected hypergraph is called *unicyclic* if $\text{bi}(H)$ is unicyclic. A connected hypergraph that is neither a hypertree nor unicyclic is called *complex*. Using the standard formula to calculate the cyclomatic number of a graph⁶ [20], we get the following (in)equalities for a connected d -uniform hypergraph H with n edges and m vertices: $(d-1) \cdot n = m - 1$ if H is a hypertree, $(d-1) \cdot n = m$ if H is unicyclic, and $(d-1) \cdot n > m$ if H is complex.

We remark that there are different notions with respect to cycles in hypergraphs. In other papers, e.g., [15, 21, 7], a hypergraph is called *acyclic* if and only if there exists a sequence of repeated deletions of edges containing at least one vertex of degree 1 that yields a hypergraph without edges. (Formally, we can arrange the edge set $E = \{e_1, \dots, e_n\}$ of the hypergraph in a sequence (e'_1, \dots, e'_n) such that

⁶The cyclomatic number of a connected graph G with m vertices and n edges is $n - m + 1$.

$e'_j - \bigcup_{s < j} e'_s \neq \emptyset$, for $1 \leq j \leq n$.) We will call this process of repeatedly removing edges incident to a vertex of degree 1 the *peeling process*, see, e.g., [55]. With respect to this definition, a hypergraph H is acyclic if and only if the 2-core of H is empty, where the 2-core of $H = (V, E)$ is the largest set $E' \subseteq E$ such that each vertex in (V, E') has minimum degree 2, disregarding isolated vertices. An acyclic hypergraph in this sense can have unicyclic and complex components according to the definition from above. In the analysis, we will point out why it is important for our work to use the concepts introduced above.

Hypergraph-Related Additions to the Framework. When working with hypergraphs, it will be helpful to allow removing single vertices from hyperedges. This motivates considering the following generalizations of the notions “peelability” and “reducibility” for hypergraph properties.

DEFINITION 4.1 (Generalized Peelability). *A hypergraph property \mathbf{A} is called **peelable** if for all $G = (V, E) \in \mathbf{A}$, $|E| \geq 1$, there exists an edge $e \in E$ such that*

1. $(V, E \setminus \{e\}) \in \mathbf{A}$ or
2. *there exists an edge $e' \subseteq e$ with $|e'| < |e|$ and $|e'| \geq 2$, where e and e' have the same label, such that $(V, (E \setminus \{e\}) \cup \{e'\}) \in \mathbf{A}$.*

DEFINITION 4.2 (Generalized Reducibility). *Let $c \in \mathbb{N}$, and let \mathbf{A} and \mathbf{B} be hypergraph properties. \mathbf{A} is called **B-2c-reducible** if for all graphs $(V, E) \in \mathbf{A}$ and sets $E^* \subseteq E$ with $|E^*| \leq 2c$ we have the following: There exists a subgraph (V, E') of (V, E) with $(V, E') \in \mathbf{B}$ such that each edge $e' \in E'$ is a subset of some $e \in E$ with the same label and for each edge $e^* \in E^*$ there exists an edge $e' \in E'$ with $e' \subseteq e^*$ and e' and e^* having the same label.*

In contrast to Definition 2.9 and Definition 2.10, we can remove vertices from edges in a single peeling or reduction step. A proof analogous to the proof of Lemma 2.11 shows that the statement of that lemma is true in the hypergraph setting as well.

4.1. Generalized Cuckoo Hashing. The obvious extension of cuckoo hashing is to use a sequence $\vec{h} = (h_1, \dots, h_d)$ of $d \geq 3$ hash functions. For a given integer $d \geq 3$ and a key set $S \subseteq U$ with $|S| = n$, our hash table consists of d tables T_1, \dots, T_d , each of size $m = O(n)$, and uses d hash functions h_1, \dots, h_d with $h_i: U \rightarrow [m]$, for $i \in \{1, \dots, d\}$. A key x must be stored in one of the cells $T_1[h_1(x)], T_2[h_2(x)], \dots$, or $T_d[h_d(x)]$. Each table cell contains at most one key. Searching and removing a key works in the obvious way. For the insertion procedure, note that evicting a key y from a table T_j leaves, in contrast to standard cuckoo hashing, $d - 1$ other choices where to put the key. To think about insertion procedures, it helps to introduce the concept of a certain directed graph. Given a set S of keys stored in a cuckoo hash table with tables T_1, \dots, T_d using \vec{h} , we define the following (directed) cuckoo allocation graph $G = (V, E)$, see, e.g., [47]: The vertices V correspond to the memory cells in T_1, \dots, T_d . The edge set E consists of all edges $(u, v) \in V \times V$ such that there exists a key $x \in S$ so that x is stored in the table cell which corresponds to vertex u (x occupies u) and v corresponds to one of the $d - 1$ other choices of key x . If $u \in V$ has out-degree 0, we call u *free*. (The table cell which corresponds to vertex u does not contain a key.) Standard methods proposed for inserting a key (see [35]) are *breadth-first search* to find a shortest eviction sequence or a *random walk* approach. In the next section, we will study two alternative insertion strategies that were suggested recently. If an insertion fails, a new sequence of hash functions is chosen and the data structure is built anew.

If the hash functions used are fully random it is now fully understood what table sizes m make it possible w.h.p. to store a key set according to the cuckoo hashing rules

for a given number of hash functions. In 2009,⁷ this case was settled independently by Dietzfelbinger *et al.* [23], Fountoulakis and Panagiotou [36], and Frieze and Melsted [40]. Later, the random walk insertion algorithm was partially analyzed by Frieze, Melsted, and Mitzenmacher [41] and Fountoulakis, Panagiotou, and Steger [37].

Here, we study the static setting in which we ask if \vec{h} allows accommodating a given key set $S \subseteq U$ in the hash table according to the cuckoo hashing rules. This is equivalent to the question whether the hypergraph $G = G(S, \vec{h})$ built from S and \vec{h} is 1-orientable or not, i.e., whether there is an injective function that maps each edge e to a vertex on e or not. If $G(S, \vec{h})$ is 1-orientable, we call \vec{h} *suitable* for S .

We now discuss some known results for random hypergraphs. As for simple random graphs [34] there is a sharp transition phenomenon for random hypergraphs [44]. When a random hypergraph with m vertices has at most $(1 - \varepsilon)m/(d(d - 1))$ edges, all components are small and all components are *hypertrees* or *unicyclic* with high probability. On the other hand, when it has at least $(1 + \varepsilon)m/(d(d - 1))$ edges, there exists one large, *complex* component. We will analyze generalized cuckoo hashing under the assumption that each table has size $m \geq (1 + \varepsilon)(d - 1)n$, for $\varepsilon > 0$. Note that this result is rather weak: The load of the hash table is at most $1/(d(d - 1))$, i.e., the more hash functions we use, the weaker the edge density bounds we get for the hash functions to provably work. At the end of this section, we will discuss whether this result can be improved or not with the methodology used here.

We will show the following theorem.

THEOREM 4.3. *Let $\varepsilon > 0, 0 < \delta < 1, d \geq 3$ be given. Assume $c \geq 2/\delta$. For $n \geq 1$, consider $m \geq (1 + \varepsilon)(d - 1)n$ and $\ell = n^\delta$. Let $S \subseteq U$ with $|S| = n$. Then for $\vec{h} = (h_1, \dots, h_d)$ chosen at random from $\mathcal{Z} = \mathcal{Z}_{\ell, m}^{c, d}$ the following holds:*

$$\Pr(\vec{h} \text{ is not suitable for } S) = O(1/n).$$

Most of this subsection is devoted to the proof of this theorem.

LEMMA 4.4. *Let H be a hypergraph. If H contains no complex component then H is 1-orientable.*

Proof. We may consider each connected component of H separately. First, observe that a hypertree and a unicyclic component always contains an edge that is incident to a vertex of degree 1.

Suppose C is such a hypertree or a unicyclic component. A 1-orientation of C is obtained via the well-known “peeling process”, see, e.g., [55]. It works by iteratively peeling edges incident to a vertex of degree 1 and orienting each edge towards such a vertex. \square

In the light of Lemma 4.4 we bound the probability of $G(S, \vec{h})$ being 1-orientable by the probability that $G(S, \vec{h})$ contains no complex component. A connected complex component of H causes $\text{bi}(G)$ to have at least two cycles. So, minimal obstruction hypergraphs that show that a hypergraph contains a complex component are very much like the obstruction graphs that showed that a graph contains more than one cycle, see Figure 2.1 on Page 14. For a clean definition of obstruction hypergraphs, we will first introduce the concept of a *strict path* in a hypergraph. A sequence (e_1, \dots, e_t) with $t \geq 1$ and $e_i \in E$, for $1 \leq i \leq t$, is a *strict path* in H if $|e_i \cap e_{i+1}| = 1$ for

⁷Technical report versions of all papers cited here were published at www.arxiv.org. We refer to the final publications.

$1 \leq i \leq t-1$ and $|e_i \cap e_j| = 0$ for $1 \leq i \leq t-2$ and $i+2 \leq j \leq t$. According to [44], a complex connected component contains a subgraph of one of the following two types:

Type 1: A strict path $e_1, \dots, e_t, t \geq 1$, and an edge f such that $|f \cap e_1| \geq 1, |f \cap e_t| \geq 1$, and $|f \cap \bigcup_{i=1}^t e_i| \geq 3$.

Type 2: A strict path $e_1, \dots, e_{t-1}, t \geq 2$, and edges f_1, f_2 such that $|f_1 \cap e_1| \geq 1, |f_2 \cap e_{t-1}| \geq 1$, and $|f_j \cap \bigcup_{i=1}^{t-1} e_i| \geq 2$, for $j \in \{1, 2\}$.

The bipartite representation of a hypergraph of Type 1 contains a cycle with a chord, the bipartite representation of a hypergraph of Type 2 contains two cycles connected by a path of some nonnegative length. We call a hypergraph H in $\mathcal{G}_{m,n}^d$ of Type 1 or Type 2 a *minimal complex obstruction hypergraph*. Let MCOG denote the set of all minimal complex obstruction hypergraphs in $\mathcal{G}_{m,n}^d$, with edges labeled by distinct elements from $\{1, \dots, n\}$. In the following, our objective is to apply Lemma 2.7, which says that

$$\Pr(N_S^{\text{MCOG}} > 0) \leq \Pr(B_S^{\text{MCOG}}) + E^*(N_S^{\text{MCOG}}). \quad (4.1)$$

Bounding $E^(N_S^{\text{MCOG}})$.* We prove the following lemma:

LEMMA 4.5. *Let $S \subseteq U$ with $|S| = n, d \geq 3$, and $\varepsilon > 0$ be given. Assume $m \geq (1 + \varepsilon)(d-1)n$. Then*

$$E^*(N_S^{\text{MCOG}}) = O(1/n).$$

Proof. From the proof of [44, Theorem 4, P. 128], we know that the number $w_1(d, t+1)$ of *unlabeled* minimal complex obstruction hypergraphs with $t+1$ edges is at most

$$w_1(d, t+1) \leq dm^d((d-1)m^{d-1})^{t-2} t^2 d^4 \left((d-1)m^{d-1} m^{d-3} + m^{2(d-2)} \right) \quad (4.2)$$

So, the number of minimal complex obstruction hypergraphs with $t+1$ edges and edge labels from $\{1, \dots, n\}$ is not larger than

$$\begin{aligned} & \left(\binom{n}{t+1} \cdot (t+1)! \right) dm^d((d-1)m^{d-1})^{t-2} \cdot t^2 d^4 \cdot \left((d-1)m^{d-1} \cdot m^{d-3} + m^{2(d-2)} \right) \\ & \leq n^{t+1} \cdot d^6 \cdot m^{(d-1)(t+1)-1} \cdot t^2 \cdot (d-1)^{t-2} \\ & \leq n^{d(t+1)-1} \cdot d^6 \cdot t^2 \cdot (1+\varepsilon)^{(d-1)(t+1)-1} \cdot (d-1)^{(d-1)(t+1)+t-3} \\ & = n^{d(t+1)-1} \cdot d^6 \cdot t^2 \cdot (1+\varepsilon)^{(d-1)(t+1)-1} \cdot (d-1)^{d(t+1)-4}. \end{aligned}$$

Let H be a labeled minimal complex obstruction hypergraph with $t+1$ edges.

Draw $t+1$ edges at random from $[m]^d$, one for each labeled edge in H . The probability that the hash values realize H is $1/m^{d(t+1)} \leq 1/((1+\varepsilon)(d-1)n)^{d(t+1)}$. So,

$$\begin{aligned} E^*(N_S^{\text{MCOG}}) & \leq \sum_{t=1}^n \frac{d^6 \cdot t^2 \cdot (1+\varepsilon)^{(d-1)(t+1)-1} \cdot (d-1)^{d(t+1)-4} \cdot n^{d(t+1)-1}}{((1+\varepsilon)(d-1)n)^{d(t+1)}} \\ & \leq \frac{d^6}{(d-1)^4 n} \cdot \sum_{t=1}^n \frac{t^2}{(1+\varepsilon)^{t-1}} = O\left(\frac{1}{n}\right). \quad \square \end{aligned}$$

Bounding $\Pr(B_S^{\text{MCOG}})$. We will now prove the following lemma.

LEMMA 4.6. *Let $S \subseteq U$ with $|S| = n, d \geq 3$, and $\varepsilon > 0$ be given. Assume $m \geq (1 + \varepsilon)(d - 1)n$. Let $\ell, c \geq 1$. Choose $\vec{h} \in \mathcal{Z}_{\ell, m}^{c, d}$ at random. Then*

$$\Pr(B_S^{\text{MCOG}}) = O\left(\frac{n}{\ell^c}\right).$$

To use our framework from Section 2, we have to find a suitable peelable hypergraph property that contains MCOG. Since minimal complex obstruction hypergraphs are path-like, we relax that notion in the following way.

DEFINITION 4.7. *Let PX be the set of all hypergraphs H from $\mathcal{G}_{m, n}^d$ that fall in one of the following categories:*

1. *H has hypergraph property MCOG.*
2. *H is a strict path.*
3. *H consists of a strict path e_1, \dots, e_t , $t \geq 1$, and an edge f such that $|f \cap (e_1 \cup e_t)| \geq 1$ and $|f \cap \bigcup_{i=1}^t e_i| \geq 2$.*

All hypergraphs in PX are extensions of paths. Note that property 3 is somewhat artificial to deal with the case that a single edge of a minimal complex obstruction hypergraph of Type 2 is removed. Obviously, MCOG is contained in PX, and property PX is peelable. We can now prove Lemma 4.6.

Proof of Lemma 4.6. We apply Lemma 2.11 which says that

$$\Pr(B_S^{\text{PX}}) \leq \frac{1}{\ell^c} \sum_{t=1}^n t^{2c} \mu_t^{\text{PX}}.$$

We start by counting unlabeled hypergraphs $G \in \text{PX}$ having exactly $t + 1$ edges. For the hypergraphs having Property 1 of Definition 4.7, we may use the bound (4.2) in the proof of Lemma 4.5. Let $w_2(d, t + 1)$ be the number of such hypergraphs which are strict paths, i.e., have Property 2 of Definition 4.7. We obtain the following bound:

$$w_2(d, t + 1) \leq dm^d \cdot ((d - 1) \cdot m^{d-1})^t.$$

Let $w_3(d, t + 1)$ be the number of hypergraphs having Property 3 of Definition 4.7. We observe that

$$w_3(d, t + 1) \leq dm^d \cdot ((d - 1) \cdot m^{d-1})^{t-1} \cdot 2d^2 \cdot t \cdot m^{d-2}.$$

So, the number of fully labeled hypergraphs having exactly $t + 1$ edges is at most

$$\begin{aligned} & \left(\binom{n}{t+1} \cdot (t+1)! \right) \cdot (w_1(d, t+1) + w_2(d, t+1) + w_3(d, t+1)) \\ & \leq n^{t+1} \cdot dm^d \cdot ((d-1) \cdot m^{d-1})^{t-2} \cdot \\ & \quad \left(d^2 m^{2(d-1)} + d^4 t m^{2(d-2)} + d^4 t^2 m^{2(d-2)} + 2d^3 t m^{2d-3} \right) \\ & \leq 4 \cdot d^5 \cdot t^2 \cdot n^{t+1} \cdot m^{(d-1)(t+1)+1} \cdot (d-1)^{t-2} \\ & \leq 4 \cdot d^5 \cdot t^2 \cdot n^{d(t+1)+1} \cdot (1+\varepsilon)^{(d-1)(t+1)+1} \cdot (d-1)^{d(t+1)-2}. \end{aligned}$$

We may thus calculate:

$$\begin{aligned} \Pr(B_S^{\text{PX}}) &\leq \frac{1}{\ell^c} \sum_{t=1}^n t^{2c} \cdot \frac{4 \cdot d^5 \cdot t^2 \cdot (1+\varepsilon)^{(d-1)(t+1)+1} \cdot (d-1)^{d(t+1)-2} \cdot n^{d(t+1)+1}}{((1+\varepsilon)(d-1)n)^{d(t+1)}} \\ &\leq \frac{n}{\ell^c} \sum_{t=1}^n \frac{4 \cdot d^5 \cdot t^{2+2c}}{(d-1)^2 \cdot (1+\varepsilon)^{t-1}} = O\left(\frac{n}{\ell^c}\right). \quad \square \end{aligned}$$

Putting Everything Together. Substituting the results of Lemma 4.5 and Lemma 4.6 into (4.1) yields

$$\Pr(N_S^{\text{MCOG}} > 0) = O\left(\frac{1}{n}\right) + O\left(\frac{n}{\ell^c}\right).$$

Theorem 4.3 follows by setting $c \geq 2/\delta$ and $\ell = n^\delta$.

Theorem 4.3 shows that given a key set of size n , if we use d tables of size at least $(1+\varepsilon)(d-1)n$ and hash functions from \mathcal{Z} there exists w.h.p. an assignment of the keys to memory cells according to the cuckoo hashing rules. Thus, the load of the hash table is smaller than $1/(d(d-1))$. In the fully random case, the load of the hash table rapidly grows towards 1, see, e.g., the table on Page 5 of [23]. For example, using 5 hash functions allows the hash table load to be ≈ 0.9924 . The approach followed in this section cannot yield such bounds for the following reason. When we look back at the proof of Lemma 4.4, we notice that it gives a stronger result: It shows that when a graph does not contain a complex component, it has an empty two-core, i.e., it does not contain a non-empty subgraph in which each vertex has minimum degree 2. It is known from random hypergraph theory that the appearance of a non-empty two-core becomes increasingly likely for d getting larger.⁸ So, we cannot rely on hypergraphs with empty two-cores to prove bounds for generalized cuckoo hashing that improve for increasing values of d .

Karoński and Łuczak showed in [44] that if a random d -uniform hypergraph has m vertices and at most $(1-\varepsilon)m/(d(d-1))$ edges, then all connected components have size $O(\log m)$ with high probability. On the other hand, if a random d -uniform hypergraph has at least $(1+\varepsilon)m/(d(d-1))$ edges, then there exists a unique connected component of size $\Theta(m)$. So, for $d \geq 3$ the analysis of general cuckoo hashing takes place in the presence of the giant component, which differs heavily from the analysis of standard cuckoo hashing. Whether or not hash class \mathcal{Z} allows suitable bounds for the general case will depend on whether or not there exist small subgraphs in the giant component which are sufficiently unlikely to occur in the fully random case.

Now that we turned our focus to hypergraphs, we again see that the analysis is made without exploiting details of the hash function construction, only using the general framework developed in Section 2 together with random graph theory.

4.2. Labeling-based Insertion Algorithms For Generalized Cuckoo Hashing. In the previous section we showed that when the tables are large enough, the hash functions allow storing S according to the cuckoo hashing rules with high probability. In this section we prove that such an assignment can be obtained (with high probability) with hash functions from \mathcal{Z} using two recently described insertion algorithms.

In the last section, we pointed out two natural insertion strategies for generalized cuckoo hashing: breadth-first search and random walk, described in [35]. Very recently,

⁸According to [53, p. 418] (see also [51]) for large d the 2-core of a random d -uniform hypergraph with m vertices and n edges is empty with high probability if m is bounded from below by $dn/\log d$.

Khosla [47] (2013) and Eppstein *et al.* [33] (2014) presented two new insertion strategies, which will be described next. In both algorithms, each table cell i in table T_j has a label (or counter) $l(j, i) \in \mathbb{N}$, where initially $l(j, i) = 0$ for all $j \in \{1, \dots, d\}$ and $i \in \{0, \dots, m-1\}$. The insertion of a key x works as follows: Both strategies find the table index

$$j = \arg \min_{j \in \{1, \dots, d\}} \{l(j, h_j(x))\}.$$

If $T_j[h_j(x)]$ is free then x is stored in this cell and the insertion terminates successfully. Otherwise, let y be the key that resides in $T_j[h_j(x)]$. Store x in $T_j[h_j(x)]$. The difference between the two algorithms is how they adjust the labeling. The algorithm of Khosla sets

$$l(j, h_j(x)) \leftarrow \min\{l(j', h_{j'}(x)) \mid j' \in (\{1, \dots, d\} \setminus \{j\})\} + 1,$$

while the algorithm of Eppstein *et al.* sets $l(j, h_j(x)) \leftarrow l(j, h_j(x)) + 1$. Now insert y in the same way. This is iterated until an empty cell is found or it is noticed that the insertion cannot be performed successfully.⁹ In Khosla's algorithm, the content of the label $l(j, i)$ is a lower bound for the minimal length of an eviction sequence that makes it possible to store a new element into $T_j[i]$ by moving other elements around [47, Proposition 1]. In the algorithm of Eppstein *et al.*, the label $l(j, i)$ contains the number of times the memory cell $T_j[i]$ has been overwritten. According to [33], it aims to minimize the number of write operations to a memory cell. Here we show that in the sparse setting with $m \geq (1 + \varepsilon)(d-1)n$, using class \mathcal{Z} the maximum label in the algorithm of Eppstein *et al.* is $\log \log n + O(1)$ with high probability and the maximum label in the algorithm of Khosla is $O(\log n)$ with high probability. This corresponds to results proved in these papers for fully random hash functions.

Our result when using hash functions from \mathcal{Z} is as follows. We only study the case that we want to insert the keys from a set S sequentially without deletions.

THEOREM 4.8. *Let $\varepsilon > 0$, $0 < \delta < 1$, $d \geq 3$ be given. Assume $c \geq 2/\delta$. For $n \geq 1$ consider $m \geq (1 + \varepsilon)(d-1)n$ and $\ell = n^\delta$. Let $S \subseteq U$ with $|S| = n$. Choose $\vec{h} \in \mathcal{Z}_{\ell, m}^{c, d}$ at random. Insert all keys from S according to \vec{h} in an arbitrary order, using the algorithm of Khosla. Then with probability $1 - O(1/n)$ (i) all key insertions are successful and (ii) $\max\{l(j, i) \mid i \in \{0, \dots, m-1\}, j \in \{1, \dots, d\}\} = O(\log n)$.*

THEOREM 4.9. *Let $\varepsilon > 0$, $0 < \delta < 1$, $d \geq 3$ be given. Assume $c \geq 2/\delta$. For $n \geq 1$ consider $m \geq (1 + \varepsilon)(d-1)n$ and $\ell = n^\delta$. Let $S \subseteq U$ with $|S| = n$. Choose $\vec{h} \in \mathcal{Z}_{\ell, m}^{c, d}$ at random. Insert all keys from S according to \vec{h} in an arbitrary order, using the algorithm of Eppstein *et al.* Then with probability $1 - O(1/n)$ (i) all key insertions are successful and (ii) $\max\{l(j, i) \mid i \in \{0, \dots, m-1\}, j \in \{1, \dots, d\}\} = \log \log n + O(1)$.*

For the analysis of both algorithms we assume that the insertion of an element fails if there exists a label of size $n+1$. (In this case, new hash functions are chosen and the data structure is built anew.) Hence, to prove Theorem 4.8 and Theorem 4.9 it suffices to show that statement (ii) holds. (An unsuccessful insertion yields a label with value $> n$.)

⁹ Neither in [33] nor in [47] it is described how this should be done in the cuckoo hashing setting. From the analysis presented there, when deletions are forbidden, one should do the following: Both algorithms have a counter MaxLabel , and if there exists a label $l(j, i) \geq \text{MaxLabel}$, then one should choose new hash functions and re-insert all items. For Khosla's algorithm, $\text{MaxLabel} = \Theta(\log n)$; for the algorithm of Eppstein *et al.*, one should set $\text{MaxLabel} = \Theta(\log \log n)$.

Analysis of Khosla's Algorithm. We first analyze the algorithm of Khosla. We remark that in our setting, Khosla's algorithm finds an assignment with high probability. (In [47, Section 2.1] Khosla gives an easy argument why her algorithm always finds an assignment when this is possible. In the previous section, we showed that such an assignment exists with probability $1 - O(1/n)$.) It remains to prove that the maximum label has size $O(\log n)$. We first introduce the notation used by Khosla in [47]. Recall the definition of the cuckoo allocation graph from the beginning of Section 4.1. Let G be a cuckoo allocation graph. Let $F_G \subseteq V$ consist of all free vertices in G . Let $d_G(u, v)$ be the distance between u and v in G . Define

$$d_G(u, F) := \min(\{d_G(u, v) \mid v \in F\} \cup \{\infty\}).$$

Now assume that the key set S is inserted in an arbitrary order. Khosla defines a *move* as every action that writes an element into a table cell. (So, the i -th insertion is decomposed into $k_i \geq 1$ moves.) The allocation graph at the end of the p -th move is denoted by $G_p = (V, E_p)$. Let M denote the number of moves necessary to insert S . (Recall that we assume that \vec{h} is suitable for S .) Khosla shows the following connection between labels and distances to a free vertex.

PROPOSITION 4.10 ([47, Proposition 1]). *For each $p \in \{0, 1, \dots, M\}$ and each $v \in V$ it holds that $d_{G_p}(v, F_{G_p}) \geq l(j, i)$, where $T_j[i]$ is the table cell that corresponds to vertex v .*

Now fix an integer $L \geq 1$. Assume that there exists an integer p with $0 \leq p \leq M$ and a vertex v such that $d(v, F_{G_p}) = L$. Let $(v = v_0, v_1, \dots, v_{L-1}, v_L)$ be a simple path p of length L in G_p such that v_L is free. Let $x_0, \dots, x_{L-1} \subseteq S$ be the keys which occupy v_0, \dots, v_{L-1} . Then the hypergraph $G(S, \vec{h})$ contains a subgraph H that corresponds to p in the obvious way.

DEFINITION 4.11. *For given integers $L \geq 1, m \geq 1, n \geq 1, d \geq 3$, let $\text{SP}(L)$ ("simple path") consist of all hypergraphs $H = (V, \{e_1, \dots, e_L\})$ in $\mathcal{G}_{m,n}^d$ with the following properties:*

1. *For all $i \in \{1, \dots, L\}$ we have that $|e_i| = 2$. (So, H is a graph.)*
2. *For all $i \in \{1, \dots, L-1\}$, $|e_i \cap e_{i+1}| = 1$.*
3. *For all $i \in \{1, \dots, L-2\}$ and $j \in \{i+2, \dots, L\}$, $|e_i \cap e_j| = 0$.*

Our goal in the following is to show that there exists a constant c such that for all $L \geq c \log n$ we have $\Pr(N_S^{\text{SP}(L)} > 0) = O(1/n)$. From Lemma 2.7 we obtain the bound

$$\Pr(N_S^{\text{SP}(L)} > 0) \leq \mathbb{E}^*(N_S^{\text{SP}(L)}) + \Pr(B_S^{\text{SP}(L)}). \quad (4.3)$$

Bounding $\mathbb{E}^(N_S^{\text{SP}(L)})$.* We show the following lemma.

LEMMA 4.12. *Let $S \subseteq U$ with $|S| = n, d \geq 3$, and $\varepsilon > 0$ be given. Consider $m \geq (d-1)(1+\varepsilon)n$. Then*

$$\mathbb{E}^*(N_S^{\text{SP}(L)}) \leq \frac{md}{(1+\varepsilon)^L}.$$

Proof. We count fully labeled hypergraphs with property $\text{SP}(L)$. Let P be an unlabeled simple path of length L . There are $d \cdot (d-1)^L$ ways to label the vertices on P with $\{1, \dots, d\}$ to fix the class of the partition they belong to. Then there are not more than m^{L+1} ways to label the vertices with labels from $[m]$. There are fewer than

n^L ways to label the edges with labels from $\{1, \dots, n\}$. Fix such a fully labeled path P' . Now draw $2L$ hash values from $[m]$ according to the labels of P' . The probability that these random choices realize P' is $1/m^{2L}$. We calculate:

$$\mathbb{E}^*(N_S^{\text{SP}(L)}) \leq \frac{d \cdot (d-1)^L \cdot m^{L+1} \cdot n^L}{m^{2L}} = \frac{m \cdot d \cdot (d-1)^L}{((d-1)(1+\varepsilon))^L} = \frac{md}{(1+\varepsilon)^L}. \quad \square$$

Bounding $\Pr(B_S^{\text{SP}(L)})$. Note that $\text{SP}(L)$ is not peelable. We relax $\text{SP}(L)$ in the obvious way and define $\text{RSP}(L) = \bigcup_{0 \leq i \leq L} \text{SP}(L)$. Graph property $\text{RSP}(L)$ is peelable.

LEMMA 4.13. *Let $S \subseteq U$ with $|S| = n$ and $d \geq 3$ be given. For an $\varepsilon > 0$, set $m \geq (1+\varepsilon)(d-1)n$. Let $\ell, c \geq 1$. Choose $\vec{h} \in \mathcal{Z}_{\ell, m}^{c, d}$ at random. Then*

$$\Pr(B_S^{\text{SP}(L)}) = O\left(\frac{n}{\ell^c}\right).$$

Proof. Since $\text{SP}(L) \subseteq \text{RSP}(L)$ and $\text{RSP}(L)$ is peelable, we may apply Lemma 2.11 and obtain the bound

$$\Pr(B_S^{\text{SP}(L)}) \leq \frac{1}{\ell^c} \cdot \sum_{t=1}^n t^{2c} \cdot \mu_t^{\text{RSP}(L)}.$$

By the definition of $\text{RSP}(L)$ and using the same counting argument as in the proof of Lemma 4.12, we calculate:

$$\Pr(B_S^{\text{SP}(L)}) \leq \frac{1}{\ell^c} \cdot \sum_{t=1}^n t^{2c} \cdot \frac{md}{(1+\varepsilon)^t} = O\left(\frac{n}{\ell^c}\right). \quad \square$$

Putting Everything Together. Plugging the results of Lemma 4.12 and Lemma 4.13 into (4.3) shows that

$$\Pr(N_S^{\text{SP}(L)} > 0) \leq \frac{md}{(1+\varepsilon)^L} + O\left(\frac{n}{\ell^c}\right).$$

Setting $L = 2 \log_{1+\varepsilon}(n)$, $\ell = n^\delta$, and $c \geq 2/\delta$ finishes the proof of Theorem 4.8.

Analysis of the Algorithm of Eppstein et al. We now analyze the algorithm of Eppstein et al. [33]. We use the *witness tree technique* to prove Theorem 4.9. This proof technique was introduced by Meyer auf der Heide, Scheideler, and Stemmann [52] in the context of shared memory simulations, and is one of the main techniques to analyze load balancing processes (see, e.g., [13, 14, 70, 67, 73]), which will be the topic of the next section.

Central to our analysis is the notion of a *witness tree for wear k* , for an integer $k \geq 1$. (Recall that in the algorithm of Eppstein et al., the label $l(j, i)$ denotes the number of times the algorithm has put a key into the cell $T_j[i]$. This is also called the *wear* of the table cell.) For given values n and m , a witness tree for wear k is a $(d-1)$ -ary tree with $k+1$ levels in which each non-leaf node is labeled with a tuple (j, i, κ) , for $1 \leq j \leq d$, $0 \leq i \leq m-1$, and $1 \leq \kappa \leq n$, and each leaf is labeled with a tuple (j, i) , $1 \leq j \leq d$ and $0 \leq i \leq m-1$. Two children of a non-leaf node v must have different first components (j -values) and, if they exist, third components (κ -values). In addition, the κ -values of a node and its children must differ.

We call a witness tree *proper* if no two different non-leaf nodes have the same labeling. Further, we say that a witness tree T can be *embedded into* $G(S, \vec{h})$ if for each non-leaf node v with label (j_0, i_0, κ) with children labeled $(j_1, i_1), \dots, (j_{d-1}, i_{d-1})$ in the first two label components in T , $h_{j_k}(x_\kappa) = i_k$, for each $0 \leq k \leq d-1$. We can think of a proper witness tree as an edge-labeled hypergraph from $\mathcal{G}_{m,n}^d$ by building from each non-leaf node labeled (j_0, i_0, κ) together with its $d-1$ children with label components $(j_1, i_1), \dots, (j_{d-1}, i_{d-1})$ a hyperedge (i'_0, \dots, i'_{d-1}) labeled “ κ ”, where i'_0, \dots, i'_{d-1} are ordered according to the j -values.

Suppose that there exists a label $l(j, i)$ with content k for an integer $k > 0$. We now argue about what must have happened that $l(j, i)$ has such a label. In parallel, we construct the witness tree for wear k . Let T be an unlabeled $(d-1)$ -ary tree with $k+1$ levels. Let y be the key residing in $T_j[i]$. Label the root of T with (j, i, κ) , where $y = x_\kappa \in S$. Then for all other choices of y in tables $T_{j'}, j' \in \{1, \dots, d\}, j' \neq j$, we have $l(j', h_{j'}(y)) \geq k-1$. (When y was written into $T_j[i]$, $l(j, i)$ was $k-1$ and this was minimal among all choices of key y . Labels are never decreased.) Let x_1, \dots, x_{d-1} be the keys in these $d-1$ other choices of y . Label the children of the root of T with the $d-1$ tuples $(j', h_{j'}(y)), 1 \leq j' \leq d, j' \neq j$, and the respective key indices. Arguing in the same way as above, we see that for each key $x_i, i \in \{1, \dots, d-1\}$, its $d-1$ other table choices must have had a label of at least $k-2$. Label the children of the node corresponding to key x_i on the second level of T with the $d-1$ other choices, for each $i \in \{1, \dots, d-1\}$. (Note that already the third level may include nodes with the same label.) Proceeding with this construction on the levels $3, \dots, k$ gives the witness tree T for wear k . By construction, this witness tree can be embedded into $G(S, \vec{h})$.

So, all we have to do to prove Theorem 4.9 is to obtain a (good enough) bound on the probability that a witness tree for wear k can be embedded into $G(S, \vec{h})$. If a witness tree is not proper, it seems difficult to calculate the probability that this tree can be embedded into $G(S, \vec{h})$, because different parts of the witness tree correspond to the same key in S , which yields dependencies among hash values. However, we know from the last section that when $G(S, \vec{h})$ is sparse enough, i.e., $m \geq (1+\varepsilon)(d-1)n$, it contains only hypertrees and unicyclic components with probability $1 - O(1/n)$. Using a basic pruning argument, Eppstein *et al.* show that this simplifies the situation in the following way.

LEMMA 4.14 ([33, Observation 2 and Observation 3]). *Let H be a hypergraph that consists only of hypertrees and unicyclic components. Suppose H contains an embedded witness tree for wear k . Then there exists a proper witness tree for wear $k-1$ that can be embedded into H .*

Let $W_{S,k}$ be the event that there exists a witness tree for wear k that can be embedded into $G(S, \vec{h})$. To prove Theorem 4.9, we have to show that for the parameter choices in Theorem 4.9 the probability that $W_{S,k}$ occurs is $O(1/n)$.

We separate the cases that $G(S, \vec{h})$ contains a complex component and that this is not so. Let $\text{PWT}(k)$ be the set of all hypergraphs in $\mathcal{G}_{m,n}^d$ that correspond to proper witness trees for wear k . Using Theorem 4.3, we may bound:

$$\begin{aligned} \Pr(W_{S,k}) &\leq \Pr(N_S^{\text{PWT}(k-1)} > 0) + \Pr(N_S^{\text{MCOG}} > 0) \\ &\leq \Pr(B_S^{\text{PWT}(k-1)}) + E^*(N_S^{\text{PWT}(k-1)}) + \Pr(N_S^{\text{MCOG}} > 0). \end{aligned} \quad (4.4)$$

The last summand on the right-hand side of this inequality is handled by Theorem 4.3, so we may concentrate on the hypergraph property $\text{PWT}(k-1)$.

Bounding $E^(N_S^{\text{PWT}(k-1)})$.* We start by proving that the expected number of proper witness trees in $G(S, \vec{h})$ is $O(1/n)$ for the parameter choices in Theorem 4.9. We use a different proof method than Eppstein *et al.* [33], because we cannot use the statement of [33, Lemma 1]. We remark here that the following analysis could be extended to obtain bounds of $O(1/n^s)$, for $s \geq 1$. However, the last summand of (4.4) is $O(1/n)$, so this does not improve the bounds for (4.4).

LEMMA 4.15. *Let $S \subseteq U$ with $|S| = n$ and $d \geq 3$ be given. For an $\varepsilon > 0$, set $m \geq (1 + \varepsilon)(d - 1)n$. Then there exists a value $k = \log \log n + \Theta(1)$ such that*

$$E^*(N_S^{\text{PWT}(k-1)}) = O\left(\frac{1}{n}\right).$$

Proof. We first obtain a bound on the number of proper witness trees for wear $k - 1$. Let T be an unlabeled $(d - 1)$ -ary tree with k levels. The number v_{k-1} of vertices of such a tree is

$$v_{k-1} = \sum_{i=0}^{k-1} (d-1)^i = \frac{(d-1)^k - 1}{d-2}.$$

For the number e_{k-1} of non-leaf nodes of such a tree, we have

$$e_{k-1} = \sum_{i=0}^{k-2} (d-1)^i = \frac{(d-1)^{k-1} - 1}{d-2}.$$

There are $n \cdot d \cdot m$ ways to label the root of T . There are not more than $n^{d-1} \cdot m^{d-1}$ ways to label the second level of the tree. Labeling the remaining levels in the same way, we see that in total there are fewer than $n^{e_{k-1}} \cdot d \cdot m^{v_{k-1}}$ proper witness trees for wear $k - 1$. Fix such a fully labeled witness tree T . Now draw $d \cdot e_{k-1} = v_{k-1} + e_{k-1} - 1$ values randomly from $[m]$ according to the labeling of the nodes in T . The probability that these values realize T is exactly $1/m^{v_{k-1} + e_{k-1} - 1}$. We obtain the following bound:

$$\begin{aligned} E^*(N_S^{\text{PWT}(k-1)}) &\leq \frac{n^{e_{k-1}} \cdot d \cdot ((1 + \varepsilon)(d - 1)n)^{v_{k-1}}}{((1 + \varepsilon)(d - 1)n)^{v_{k-1} + e_{k-1} - 1}} = \frac{n \cdot d}{((1 + \varepsilon)(d - 1))^{e_{k-1} - 1}} \\ &\leq \frac{n \cdot d}{((1 + \varepsilon)(d - 1))^{(d-1)^{k-2}}}, \end{aligned}$$

which is $O(1/n)$ for $k = \log \log n + \Theta(1)$. \square

Bounding $\Pr(B_S^{\text{PWT}(k-1)})$. We first relax the notion of a witness tree in the following way.

DEFINITION 4.16. *Let $\text{RWT}(k - 1)$ (relaxed witness trees) be the set of all hypergraphs which can be obtained in the following way:*

1. *Let $T \in \text{PWT}(k')$ be an arbitrary proper witness tree for wear k' , with $k' \leq k - 1$. Let ℓ denote the number of nodes on level $k' - 1$, i.e., the level prior to the leaf level of T .*
2. *Arbitrarily choose $\ell' \in \mathbb{N}$ with $\ell' \leq \ell - 1$.*
3. *Choose $\kappa = \lfloor \ell' / (d - 1) \rfloor$ arbitrary distinct non-leaf nodes on level $k' - 2$. For each such node, remove all its children together with their $d - 1$ children from T . Then remove from a group of $d - 1$ siblings on level $k' - 1$ the $\ell' - (d - 1) \cdot \kappa$ siblings with the largest j -values together with their leaves.*

Note that $\text{RWT}(k-1)$ is a peelable hypergraph property, for we can iteratively remove non-leaf nodes that correspond to edges in the hypergraph until the whole leaf level is removed. Removing these nodes as described in the third property makes sure that there exists at most one non-leaf node at level $k'-2$ that has fewer than $d-1$ children. Also, it is clear what the first components in the labeling of the children of this node are. Removing nodes in a more arbitrary fashion would give more labeling choices and thus more trees with property $\text{RWT}(k-1)$.

LEMMA 4.17. *Let $S \subseteq U$ with $|S| = n$ and $d \geq 3$ be given. For an $\varepsilon > 0$, set $m \geq (1 + \varepsilon)(d-1)n$. Let $\ell, c \geq 1$. Choose $\vec{h} \in \mathcal{Z}_{\ell, m}^{c, d}$ at random. Then*

$$\Pr(B_S^{\text{RWT}(k-1)}) = O\left(\frac{n}{\ell^c}\right).$$

Proof. We apply Lemma 2.11, which says that

$$\Pr(B_S^{\text{RWT}(k-1)}) \leq \frac{1}{\ell^c} \cdot \sum_{t=2}^n t^{2c} \mu_t^{\text{RWT}(k-1)}.$$

Using the same line of argument as in the bound for $E^*(N_S^{\text{PWT}(k-1)})$, the expected number of witness trees with property $\text{RWT}(k-1)$ with exactly t edges, i.e., exactly t non-leaf nodes, is at most $n \cdot d \cdot ((1 + \varepsilon)(d-1))^{t-1}$. We calculate:

$$\Pr(B_S^{\text{RWT}(k-1)}) = \frac{1}{\ell^c} \cdot \sum_{t=1}^n t^{2c} \frac{n \cdot d}{((1 + \varepsilon)(d-1))^{t-1}} = O\left(\frac{n}{\ell^c}\right). \quad \square$$

Putting Everything Together. Using (4.4) and Lemmas 4.15 and 4.17, we conclude that

$$\begin{aligned} \Pr(W_{S, k}) &\leq \Pr(B_S^{\text{PWT}(k-1)}) + E^*(N_S^{\text{PWT}(k-1)}) + \Pr(N_S^{\text{MCOG}} > 0) \\ &= O(1/n) + O(n/\ell^c). \end{aligned}$$

Theorem 4.9 follows for $\ell = n^\delta$ and $c \geq 2/\delta$.

With respect to the algorithm of Eppstein *et al.*, our result shows that n insertions take time $O(n \log \log n)$ with high probability when using hash functions from \mathcal{Z} . With an analogous argument to the one given by Khosla in [47], the algorithm of Eppstein *et al.* of course finds an assignment of the keys whenever this is possible. However, the bound of $O(\log \log n)$ on the maximum label is only known for $m \geq (1 + \varepsilon)(d-1)n$ and $d \geq 3$, even in the fully random case. Extending the analysis on the maximum label size to denser hypergraphs is an interesting open question.

4.3. Load Balancing. In this section we apply hash class \mathcal{Z} in the area of load balancing schemes. In the discussion at the end of this section, we will present a link of our results w.r.t. load balancing to the space utilization of generalized cuckoo hashing in which each memory cell can hold $\kappa \geq 1$ items.

In randomized load balancing we want to allocate a set of jobs J to a set of machines M such that a condition, e.g., there exists no machine with “high” load, is satisfied with high probability. To be consistent with the notation used in our framework and previous applications, S will denote the set of jobs, and the machines will be numbered $1, \dots, m$. In this section we assume $|S| = n = m$, i.e., we allocate n jobs to n machines.

We use the following approach to load balancing: For an integer $d \geq 2$, we split the n machines into groups of size n/d each. For simplicity, we assume that d divides n . Now a job chooses d candidate machines by choosing exactly one machine from each group. This can be modeled by using d hash functions h_1, \dots, h_d with $h_i: S \rightarrow [n/d], 1 \leq i \leq d$, such that machine $h_i(j)$ is the candidate machine in group i of job j .

In load balancing schemes, the arrival of jobs has been split into two models: parallel and sequential arrival. We will focus on parallel job arrivals and come back to the sequential case at the end of this section.

In the parallel arrival model, all jobs arrive at the same time. They communicate with the machines in synchronous rounds. In these rounds, decisions on the allocations of jobs to machines are made. The τ -collision protocol is one algorithm to find such an assignment. This protocol was studied in the context of distributed memory machines by Dietzfelbinger and Meyer auf der Heide [27]. As a method for load balancing the allocation algorithm was analyzed by Stemmann in [70]. The τ -collision protocol works in the following way: First, each job chooses one candidate machine from each of the $d \geq 2$ groups. Then the following steps are repeated until all jobs are assigned to machines:

1. Synchronously and in parallel, each unassigned job sends an allocation request to each of its candidate machines.
2. Synchronously and in parallel, each machine sends an acknowledgement to all requesting jobs if and only if it got at most τ allocation requests in this round. Otherwise, it does not react.
3. Each job that gets an acknowledgement is assigned to one of the machines that has sent an acknowledgement. Ties are broken arbitrarily.

Note that the number of rounds is not bounded. However, in [27] and [70] it was shown that w.h.p. the τ -collision protocol will terminate after a small number of rounds, if suitable hash classes are used. We will show that this also holds when class \mathcal{Z} is used.

There exist several analysis techniques for load balancing, e.g., layered induction, fluid limit models and witness trees [64]. We will focus on the witness tree technique. We use the variant studied by Schickinger and Steger in [67] in connection with hash class \mathcal{Z} . The main contribution of [67] is that it provides a unified analysis for several load balancing algorithms. This allows us to show that hash class \mathcal{Z} is suitable in all of these situations as well, with only little additional work.

At the core of the analysis in [67] is the so-called *allocation graph*. In our setting, where each job chooses exactly one candidate machine in each of the d groups, the allocation graph is a bipartite graph $G = ([n], [n], E)$, where the jobs are on the left side of the bipartition, and the machines are on the right side, split into groups of size n/d . Each job vertex is adjacent to its d candidate machines. As already discussed in Section 4.1, the allocation graph is equivalent to the hypergraph $G(S, \vec{h})$. Recall that we refer to the bipartite representation of a hypergraph $G = (V, E)$ as $\text{bi}(V, E)$. We call the vertices on the left side *job vertices* and the vertices on the right side *machine vertices*.

If a machine has high load we can find a subgraph in the allocation graph that shows the chain of events in the allocation process that led to this situation, hence “witnessing” the high load of this machine. (This is similar to the wear of a table cell in the algorithm of Eppstein *et al.* [33] in the previous section.) Such witness trees might differ greatly in structure, depending on the load balancing scheme.

In short, the approach of Schickinger and Steger works as follows.¹⁰

1. They show that high load leads to the existence of a “witness graph” and describe the properties of such a graph for a given load balancing scheme.
2. For their analysis to succeed they demand that the witness graph from above is a tree in the standard sense. They show that with high probability a witness graph can be turned into a cycle-free witness tree by removing a small number of edges at the root.
3. They show that it is unlikely that the allocation graph contains such a witness tree.

We will give a detailed description of this approach after stating the main result of this section.

The following theorem represents one selected result from [67], replacing the full randomness assumption with hash functions from \mathcal{Z} to choose candidate machines for jobs. We simplify the theorem by omitting the exact parameter choices calculated in [67]. All the other examples considered in [67] can be analyzed in an analogous way, resulting in corresponding theorems. We discuss this claim further in the discussion part of this section.

THEOREM 4.18. *For each constant $\alpha > 0, d \geq 2$, there exist constants $\beta, c > 0$ (depending on α and d), such that for each t with $2 \leq t \leq (1/\beta) \ln \ln n, \ell = n^{1/2}$ and $\vec{h} = (h_1, \dots, h_d) \in \mathcal{Z}_{\ell, n}^{c, d}$, the τ -collision protocol described above with threshold $\tau = O((\ln n)/(\ln \ln n))^{1/(t-2)/(d-1)}$ finishes after t rounds with probability $1 - O(n^{-\alpha})$.*

We will now analyze the τ -collision protocol using hash functions from class \mathcal{Z} . Most importantly, we have to describe the probability of the event that the τ -collision protocol does not terminate after t rounds in the form of a hypergraph property. To achieve this, we start by describing the structure of witness trees.

In the setting of the τ -collision protocol in parallel arrival, a witness tree has the following structure. Using the notation of [67], a machine is *active in round t* if there exists at least one job that sends a request to this machine in round t . If no such job exists, the machine is *inactive in round t* . Assume that after round t the collision protocol has not yet terminated. Then there exists a machine y that is active in round t and that received more than τ allocation requests. Arbitrarily choose τ of these requests. These requests were sent by τ unallocated jobs in round t . The vertex that corresponds to machine y is the root of the witness tree, the τ job vertices are its children. In round t , each of the τ unallocated jobs sent allocation requests to $d - 1$ other machines. The corresponding machine vertices are the children of each of the τ job vertices in the witness tree. By definition, the machines that correspond to these machine vertices are also active in round t , and so they were active in round $t - 1$ as well. So, there are $\tau \cdot (d - 1)$ machine vertices that correspond to machines that are active and receive requests from one of the jobs on level t in round $t - 1$. We must be aware that among these machine vertices the same machine might appear more than once, because unallocated jobs may have chosen the same candidate machine. So, there may exist vertices in the witness tree that correspond to the same machine. For all these $\tau \cdot (d - 1)$ machines the same argument holds in round $t - 1$. Proceeding with the construction for rounds $t - 2, t - 3, \dots, 1$, we build the *witness tree T_t with root y* . It exhibits a regular recursive structure, depicted abstractly in Figure 4.1. Note that

¹⁰This approach has a lot in common with our analysis of insertion algorithms for generalized cuckoo hashing. However, the analysis will be much more complicated here, since the hypergraph $G(S, \vec{h})$ has exactly as many vertices as edges.

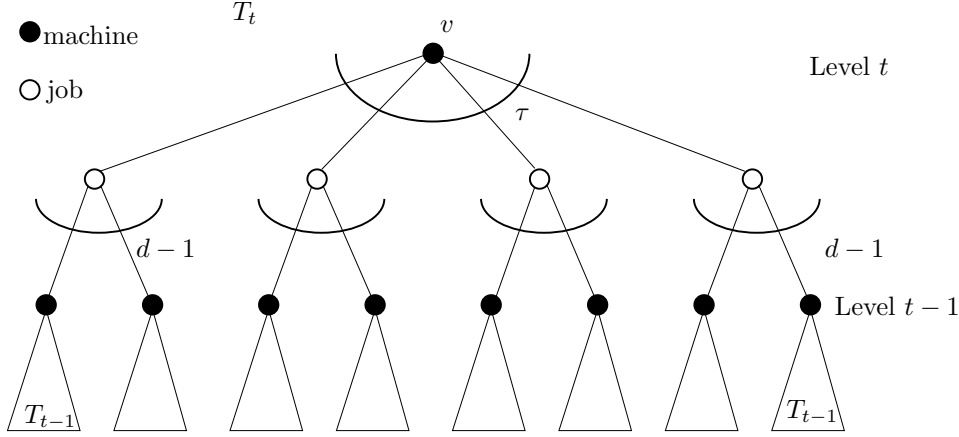


FIG. 4.1. Structure of a witness tree T_t with root v after t rounds if the τ -collision protocol with d candidate machines has not yet terminated.

all leaves, i.e., vertices on level 0, correspond to machine vertices, since no allocation requests are sent in round 0.

As we have seen, such regular witness trees do not need to be subgraphs of the allocation graph since two vertices of a witness tree might be embedded to the same vertex. Hence, the witness tree is “folded together” to a subgraph in the allocation graph. In the embedding of a witness tree as a subgraph of the allocation graph, edges do not occur independently and the analysis becomes difficult, even in the fully random case.

Schickinger and Steger found the following way to analyze this situation. For a connected undirected graph $G = (V, E)$ a *shortest path tree* T rooted at node $s \in V$ is a tree in G in which the unique paths from s to all other vertices are shortest paths in G . (Such a tree can be obtained by starting a breadth-first search in G from s .) Schickinger and Steger introduced the notion of a *multicycle* that describes an “almost tree-like” graph.

DEFINITION 4.19. Let $k, t \geq 1$. Let $G = (V, E)$ be an undirected graph. Let $s \in V$. A (k, t) -multicycle of depth at most t at node s in G is a connected subgraph $G' = (V', E')$ of G together with a shortest path tree (V', T') of G' rooted at node s with the following properties:

1. G' includes vertex s .
2. G' has cyclomatic number k (cf. Section 3.1).
3. For each vertex v in T' , the distance between s and v in T' is at most t .
4. Each leaf in T' is incident to an edge in G' that is not in T' .

Multicycles will be used to reason in the following way: When G does not contain a certain (k, t) -multicycle at node s as a subgraph, removing only a few edges in G incident to node s makes the neighborhood that includes all vertices of distance at most t of s in the remaining graph acyclic. As we shall see, the proof that this is possible will be quite easy when using shortest path trees.

One easily checks that a (k, t) -multicycle M with m vertices and n edges satisfies $n = m + k - 1$, because the cyclomatic number of a connected graph is exactly $n - m + 1$ [20]. Furthermore, it has at most $2kt + 1$ vertices, because there can be at most $2k$ leaves that each have distance at most t from s , and all vertices of the spanning tree

lie on the unique paths from s to the leaves. We will later see that for the parameters given in Theorem 4.18, a (k, t) -multicycle is with high probability not a subgraph of the allocation graph.

LEMMA 4.20 ([67, Lemma 2]). *Let $k, t \geq 1$. Assume that a graph $G = (V, E)$ contains no (k', t) -multicycle, for $k' > k$. Furthermore, given a vertex $v \in V$, consider the induced subgraph $H = (V', E')$ of G that contains all vertices $w \in V$ with distance at most t from v in G . Then we can remove at most $2k$ edges incident to v in H to get a graph H^* such that the connected component of v in H^* is a tree.*

In the light of this lemma, we set $\tau \geq 2k + 1$. Then we know that if the allocation graph contains a witness tree after t rounds, it contains a (k, t) -multicycle or a regular witness tree T_{t-1} . This observation motivates considering the following hypergraph property:

DEFINITION 4.21. *Let $k, t \in \mathbb{N}$. Then $\text{MCWT}(k, t) \subseteq \mathcal{G}_{n/d, n}^d$ is the set of all hypergraphs H such that $\text{bi}(H)$ forms either a (k, t) -multicycle or a witness tree T_{t-1} .*

If we use hash class \mathcal{Z} and set $\tau \geq 2k + 1$, for a set S of jobs, we have, by the discussion above:

$$\Pr(\text{the } \tau\text{-collision protocol does not terminate after } t \text{ rounds}) \leq \Pr\left(N_S^{\text{MCWT}(k, t)} > 0\right). \quad (4.5)$$

By Lemma 2.7 we may bound the probability on the right-hand side of (4.5) by

$$\Pr\left(N_S^{\text{MCWT}(k, t)} > 0\right) \leq \Pr\left(B_S^{\text{MCWT}(k, t)}\right) + \mathbb{E}^*\left(N_S^{\text{MCWT}(k, t)}\right). \quad (4.6)$$

Bounding $\mathbb{E}^(N_S^{\text{MCWT}(k, t)})$.* We first bound the expected number of subgraphs that form multicycles or witness trees when the hash functions are fully random. The following lemma is equivalent to Theorem 1 in [67]. However, our parameter choices are slightly different because in [67] each of the d candidate machines is chosen from the set $[n]$, while we split the n machines into d groups of size n/d . The proof of the lemma can be found in Appendix A.

LEMMA 4.22. *Let $\alpha \geq 1$ and $d \geq 2$. Set $\beta = 2d(\alpha + \ln d + 3/2)$ and $k = \alpha + 2$. Consider t with $2 \leq t \leq (1/\beta) \ln \ln n$. Let*

$$\tau = \max\left\{\frac{1}{d-1} \left(\frac{\beta t \ln n}{\ln \ln n}\right)^{\frac{1}{t-2}}, d^{d+1}e^d + 1, 2k + 1\right\}.$$

Then

$$\mathbb{E}^*\left(N_S^{\text{MCWT}(k, t)}\right) = O(n^{-\alpha}).$$

Bounding $\Pr(B_S^{\text{MCWT}(k, t)})$. To apply Lemma 2.11, we need a peelable hypergraph property that contains $\text{MCWT}(k, t)$. We will first calculate the size of witness trees to see that they are small for the parameter settings given in Theorem 4.18.

The Size of Witness Trees. Let T_t be a witness tree after t rounds. As above, the number of job vertices j_t in T_t is given by

$$j_t = \frac{\tau^t(d-1)^{t-1} - \tau}{\tau(d-1) - 1}.$$

We bound the size of the witness tree.

LEMMA 4.23. *If α, d, β, k, t , and τ are as in Lemma 4.22, we have $j_t < \log n$.*

Proof. Observe the following upper bound for the number of jobs in a witness tree after t rounds:

$$j_t = \frac{\tau^t(d-1)^{t-1} - \tau}{\tau(d-1) - 1} \leq \frac{\tau(\tau(d-1))^{t-1}}{2\tau - 1} \leq (\tau(d-1))^{t-1}.$$

Now observe that for $\tau \in \{d^{d+1}e^d + 1, 2k + 1\}$ we have

$$(\tau(d-1))^{t-1} \leq (\tau(d-1))^{\frac{1}{\beta} \ln \ln n} \leq (\ln n)^{\frac{\ln \tau + \ln d}{\beta}} \leq \ln n,$$

since $\frac{\ln \tau}{\beta} \leq 1$ for the two constant choices for τ in Lemma 4.22. Furthermore, for $\tau = (\beta t(\ln n) / \ln \ln n)^{\frac{1}{t-2}} / (d-1)$ we have

$$((d-1)\tau)^{t-1} \leq \frac{\beta t \ln n}{\ln \ln n} \leq \ln n,$$

and hence $j_t \leq \ln n < \log n$. \square

A (k, t) -multicycle has at most $2kt + k - 1$ edges, hence such multicycles are smaller than witness trees for $t = O(\log \log n)$ and a constant $k \geq 1$.

A Peelable Hypergraph Property. To apply Lemma 2.11, we have to find a peelable hypergraph property that contains all subgraphs that have property MCWT(k, t) (multicycles or witness trees for $t - 1$ rounds). Since we know from above that witness trees and multicycles are contained in small connected subgraphs of the hypergraph $G(S, \vec{h})$, we will use the following hypergraph property.

DEFINITION 4.24. *Let $K > 0$ and $d \geq 2$ be constants. Let $C_{\text{small}}(K, d)$ contain all connected d -partite hypergraphs $(V, E) \in \mathcal{G}_{n/d, n}^d$ with $|E| \leq K \log n$ disregarding isolated vertices.*

The following central lemma shows how we can bound the failure term of \mathcal{Z} .

LEMMA 4.25. *Let $K > 0$, $c \geq 1$, $\ell \geq 1$, and $d \geq 2$ be constants. Let S be the set of jobs with $|S| = n$. Then*

$$\Pr(B_S^{\text{MCWT}(k, t)}) \leq \Pr(B_S^{C_{\text{small}}(K, d)}) = O\left(\frac{n^{K(d+1) \log d + 2}}{\ell^c}\right).$$

For proving this bound, we need the following auxiliary hypergraph property. Note that some of the considered graphs are not d -uniform.

DEFINITION 4.26. *Let $K > 0$, $\ell \geq 1$, and $d \geq 2$ be constants. Let $n \geq 1$ be given. Then $\text{HT}(K, d, \ell)$ (“hypertree”) is the set of all d -partite hypergraphs $G = (V, E)$ in $\mathcal{G}_{n/d, n}^d$ with $|E| \leq K \log n$ for which $\text{bi}(G)$ (disregarding isolated vertices) is a tree, has at most ℓ leaf edges and has leaves only on the left (job) side.*

We will now establish the following connection between $C_{\text{small}}(K, d)$ and $\text{HT}(K, d, \ell)$.

LEMMA 4.27. *Let $K > 0$, $d \geq 2$ and $c \geq 1$ be constants. Then $C_{\text{small}}(K, d)$ is $\text{HT}(K, d, 2c)$ - $2c$ -reducible, cf. Definition 4.2.*

Proof. Assume $G = (V, E) \in C_{\text{small}}(K, d)$. Arbitrarily choose $E^* \subseteq E$ with $|E^*| \leq 2c$. We have to show that there exists an edge set E' such that $(V, E') \in \text{HT}(K, d, 2c)$, (V, E') is a subgraph of (V, E) , and for each edge $e^* \in E^*$ there exists an edge $e' \in E'$ such that $e' \subseteq e^*$ and e' and e^* have the same label.

Identify an arbitrary spanning tree T in $\text{bi}(G)$. Now repeatedly remove leaf vertices with their incident edges, as long as these leaf vertices do not correspond to edges from

E^* . Denote the resulting tree by T' . In the hypergraph representation, T' satisfies all properties from above. \square

From Lemma 2.11 it follows that we have

$$\Pr\left(B_S^{\text{MCWT}(k,t)}\right) \leq \Pr\left(B_S^{\mathcal{C}_{\text{small}}(K,d)}\right) \leq \ell^{-c} \cdot \sum_{t=2}^n t^{2c} \mu_t^{\text{HT}(K,d,2c)}.$$

LEMMA 4.28. *Let $K > 0, d \geq 2$, and $c \geq 1$ be constants. If $t \leq K \cdot \log n$, then*

$$\mu_t^{\text{HT}(K,d,2c)} \leq t^{O(1)} \cdot n \cdot d^{(d+1)t}.$$

For $t > K \log n$ it holds that $\mu_t^{\text{HT}(K,d,2c)} = 0$.

Proof. It is trivial that $\mu_t^{\text{HT}(K,d,2c)} = 0$ for $t > K \log n$, since a hypergraph with more than $K \log n$ edges contains too many edges to have property $\text{HT}(K, d, 2c)$.

Now suppose $t \leq K \log n$. We first count labeled hypergraphs having property $\text{HT}(K, d, 2c)$ consisting of t job vertices and z edges, for some fixed $z \in \{t, \dots, dt\}$, in the *bipartite representation*. (Note that hypergraphs with property $\text{HT}(K, d, 2c)$ may not be d -uniform. Thus, in the bipartite representation not all vertices on the left side have d neighbors on the right side.)

There are at most $z^{O(2c)} = z^{O(1)}$ unlabeled trees with z edges and at most $2c$ leaf edges (Lemma 3.2). Fix one such tree T . There are not more than n^t ways to label the job vertices of T , and there are at most d^z ways to assign each edge a label from $\{1, \dots, d\}$. Once these labels are fixed, there are at most $(n/d)^{z+1-t}$ ways to assign the right vertices to machines. Fix such a fully labeled tree T' .

Now draw z hash values at random from $[n/d]$ and build a graph according to these hash values and the labels of T' . The probability that these random choices realize T' is exactly $1/(n/d)^z$. Thus we may estimate:

$$\begin{aligned} \mu_t^{\text{HT}(K,d,2c)} &\leq \sum_{z=t}^{dt} \frac{(n/d)^{z+1-t} \cdot z^{O(1)} \cdot n^t \cdot d^z}{(n/d)^z} = \sum_{z=t}^{dt} z^{O(1)} \cdot n \cdot d^{z-1+t} \\ &< dt \cdot (dt)^{O(1)} \cdot n \cdot d^{(d+1)t} = t^{O(1)} \cdot n \cdot d^{(d+1)t}. \quad \square \end{aligned}$$

We can now proceed with the proof our main lemma.

Proof of Lemma 4.25. By Lemma 2.11, we know that

$$\Pr\left(B_S^{\text{MCWT}(k,t)}\right) \leq \ell^{-c} \cdot \sum_{t=2}^n t^{2c} \mu_t^{\text{HT}(K,d,2c)}.$$

Applying the result of Lemma 4.28, we calculate

$$\begin{aligned} \Pr\left(B_S^{\text{MCWT}(k,t)}\right) &\leq \ell^{-c} \cdot \sum_{t=2}^{K \log n} t^{2c} t^{O(1)} \cdot n \cdot d^{(d+1)t} = n \cdot \ell^{-c} \cdot (K \log n)^{O(1)} \cdot d^{(d+1)K \log n} \\ &= O(n^2 \cdot \ell^{-c}) \cdot d^{K(d+1) \log n} = O(n^{K(d+1) \log d + 2} \cdot \ell^{-c}). \quad \square \end{aligned}$$

Putting Everything Together. The previous lemmas allow us to complete the proof of the main theorem.

Proof of Theorem 4.18.

We plug the results of Lemma 4.22 and Lemma 4.25 into (4.6) and get

$$\Pr\left(N_S^{\text{MCWT}(k,t)} > 0\right) \leq O\left(\frac{n^{K(d+1)\log d+2}}{\ell^c}\right) + O\left(\frac{1}{n^\alpha}\right).$$

For the case of parallel arrival with $d \geq 2$ hash functions, we calculated that witness trees do not have more than $\log n$ edges (Lemma 4.23). So, we set $K = 1$. Setting $\ell = n^{1/2}$ and $c = 2(2 + \alpha + (d+1)\log d)$ finishes the proof of the theorem. \square

Discussion on Load Balancing. We remark that the graph property $\mathcal{C}_{\text{small}}(K, d)$ provides a very general result on the failure probability of \mathcal{Z} on hypergraphs $G(S, \vec{h})$. It can be applied for all results from [67]. We will exemplify this statement by discussing what needs to be done to show that \mathcal{Z} works in the setting of Voecking’s “Always-Go-Left” sequential allocation algorithm [73]. By specifying explicitly how to break ties (always allocate the job to the “left-most” machine), Voecking’s algorithm decreases the maximum bin load (w.h.p.) in sequential load balancing with $d \geq 2$ hash functions from $\ln \ln n / \ln d + O(1)$ (arbitrary tie-breaking) [4] to $\ln \ln n / (d \cdot \ln \Phi_d) + O(1)$, which is an exponential improvement in d . Here Φ_d is defined as follows. Let $F_d(j) = 0$ for $j \leq 0$ and $F_d(1) = 1$. For $j \geq 2$, $F_d(j) = \sum_{i=1}^d F_d(j-i)$. (This is a generalization of the Fibonacci numbers.) Then $\Phi_d = \lim_{j \rightarrow \infty} F_d(j)^{1/j}$. It holds that Φ_d is a constant with $1.61 \leq \Phi_d \leq 2$, see [73]. (We refer to [67, Section 5.2] and [73] for details about the description of the algorithm.) In the unified witness tree approach of Schickinger and Steger, the main difference between the analysis of parallel arrivals and the sequential algorithm of Voecking is in the definition of the witness tree. Here, the analysis in [67] also assumes that the machines are split into d groups of size n/d . This means that we can just re-use their analysis in the fully random case. For bounding the failure term of hash class \mathcal{Z} , we have to show that the witness trees in the case of Voecking’s “Go-Left” algorithm (see [67, Fig. 6]) have at most $O(\log n)$ jobs, i.e., that they are contained in small connected components. Otherwise, we cannot apply Lemma 4.25.

According to [67, Page 84], the number of job vertices j_ℓ in a witness tree for a bin with load ℓ is bounded by

$$j_\ell \leq 4h_\ell + 1, \quad (4.7)$$

where h_ℓ is the number of leaves in the witness tree. Following Voecking [73], Schickinger and Steger show that setting ℓ as large as $\ln \ln n / (d \ln \Phi_d) + O(1)$ is sufficient to bound the expected number of witness trees by $O(n^{-\alpha})$. Such witness trees have only $O(\log n)$ many job nodes.

LEMMA 4.29. *Let $\alpha > 0$ and $\ell = \log_{\Phi_d}(4 \log n^\alpha)/d$. Then $j_\ell \leq 33\alpha \log n$.*

Proof. It holds $h_\ell = F_d(d \cdot \ell + 1)$, see [67, Page 84], and $F_d(d \cdot \ell + 1) \leq \Phi_d^{d \cdot \ell + 1}$, since $F_d(j)^{1/j}$ is monotonically increasing. We obtain the bound

$$\begin{aligned} j_\ell &\leq 4 \cdot h_\ell + 1 \leq 4 \cdot \Phi_d^{d \cdot \ell + 1} + 1 \leq 4 \cdot \Phi_d^{\log_{\Phi_d}(4 \log n^\alpha) + 1} + 1 \\ &= 16 \cdot \Phi_d \cdot \alpha \log n + 1 \leq 33\alpha \log n, \end{aligned}$$

using $\Phi_d \leq 2$ and assuming $\alpha \log n \geq 1$. \square

Thus, we know that a witness tree in the setting of Voecking’s algorithm is contained in a connected hypergraph with at most $33\alpha \log n$ edges. Thus, we may apply Lemma 4.25 in the same way as we did for parallel arrival. The result is that for given $\alpha > 0$ we can choose $(h_1, \dots, h_d) \in \mathcal{Z}_{\ell, n}^{c, d}$ with $\ell = n^\delta$, $0 < \delta < 1$, and $c \geq (33\alpha(d+1)\log d + 2 + \alpha)/\delta$ and know that the maximum load is $(\ln \ln n)/(d \cdot$

$\kappa+1 \setminus d$	3	4	5	6	7	8
2	0.818	0.772	0.702	0.637	0.582	0.535
3	0.776	0.667	0.579	0.511	0.457	0.414
4	0.725	0.604	0.515	0.450	0.399	0.359
5	0.687	0.562	0.476	0.412	0.364	0.327
6	0.658	0.533	0.448	0.387	0.341	0.305

TABLE 4.1

Space utilization thresholds for generalized cuckoo hashing with $d \geq 3$ hash functions and $\kappa + 1$ keys per cell, for $\kappa \geq 1$, based on the non-existence of the $(\kappa + 1)$ -core. Each table cell gives the maximal space utilization achievable for the specific pair $(d, \kappa + 1)$. These values have been obtained using Maple[®] to evaluate the formula from Theorem 1 of [55].

$\ln \Phi_d) + O(1)$ with probability $1 - O(1/n^\alpha)$. So, our general analysis using small connected hypergraphs makes it very easy to show that hash class \mathcal{Z} suffices to run a specific algorithm with load guarantees.

When we are interested in making the parameters for setting up a hash function as small as possible, one should take care when bounding the constants in the logarithmic bound on the number of edges in the connected hypergraphs. (According to [67], (4.7) can be improved by a more careful argumentation.) More promising is a direct approach to witness trees, as we did in the analysis of the algorithm of Eppstein *et al.* in the previous subsection, i.e., directly peeling the witness tree. Using such an approach, Woelfel showed in [77, Theorem 2.1 and its discussion] that smaller parameters for the hash functions \mathcal{Z} are sufficient to run Voecking’s algorithm.

Application to Generalized Cuckoo Hashing. We further remark that the analysis of the τ -collision protocol makes it possible to analyze the space utilization of generalized cuckoo hashing using $d \geq 2$ hash functions and buckets which hold up to $\kappa \geq 2$ keys in each table cell, as proposed by Dietzfelbinger and Weidling in [31]. Obviously, a suitable assignment of keys to table cells is equivalent to a κ -orientation of $G(S, \vec{h})$. It is well-known that any graph that has an empty $(\kappa + 1)$ -core, i.e., that has no subgraph in which all vertices have degree at least $\kappa + 1$, has a κ -orientation, see, e.g., [18] and the references therein. The $(\kappa + 1)$ -core of a graph can be obtained by repeatedly removing vertices with degree at most κ and their incident hyperedges. The precise study of this process is due to Molloy [55]. The τ -collision protocol is the parallel variant of this process, where in each round all vertices with degree at most τ are removed with their incident edges. (In the fully random case, properties of this process were recently studied by Jiang, Mitzenmacher, and Thaler in [43].) In terms of orientability, Theorem 4.18 with the exact parameter choices from Lemma 4.22 shows that for $\tau = \max\{e^\beta, d^{d+1}e^d + 1, 2k + 1\}$ there exists (w.h.p.) an assignment of the n keys to n memory cells when each cell can hold τ keys. (This is equivalent to a hash table load of $1/\tau$.) It is open to find good space bounds for generalized cuckoo hashing using this approach. However, we think that it suffers from the same general problem as the analysis for generalized cuckoo hashing with $d \geq 3$ hash functions and one key per table cell: Since the analysis builds upon a process which requires an empty $(\kappa + 1)$ -core in the hypergraph to succeed, space utilization seems to decrease for d and κ getting larger. Table 4.1 contains space utilization bounds for static generalized cuckoo hashing with $d \geq 3$ hash functions and κ elements per table cell when the assignment is obtained via a process that requires the $(\kappa + 1)$ -core to be empty. These calculations clearly support the conjecture that space utilization decreases for larger values of d and κ .

5. A Generalized Version of the Hash Class. In this short section we present a generalized version of our hash class that uses arbitrary κ -wise independent hash classes as building blocks.

5.1. The Generalized Hash Class. The following definition is a generalization of Definition 2.3 to functions with higher degrees of independence than two.

DEFINITION 5.1. *Let $c \geq 1$, $d \geq 2$, and $\kappa \geq 2$. For integers $m, \ell \geq 1$, and given $f_1, \dots, f_d: U \rightarrow [m]$, $g_1, \dots, g_c: U \rightarrow [\ell]$, and d two-dimensional tables $z^{(i)}[1..c, 0..\ell - 1]$ with elements from $[m]$ for $i \in \{1, \dots, d\}$, we let $\vec{h} = (h_1, \dots, h_d) = (h_1, \dots, h_d) \langle f_1, \dots, f_d, g_1, \dots, g_c, z^{(1)}, \dots, z^{(d)} \rangle$, where*

$$h_i(x) = \left(f_i(x) + \sum_{1 \leq j \leq c} z^{(i)}[j, g_j(x)] \right) \bmod m, \text{ for } x \in U, i \in \{1, \dots, d\}.$$

Let $\mathcal{H}_m^\kappa [\mathcal{H}_\ell^\kappa]$ be an arbitrary κ -wise independent hash class with functions from U to $[m]$ [from U to $[\ell]$]. Then $\mathcal{Z}_{\ell, m}^{c, d, \kappa}(\mathcal{H}_\ell^\kappa, \mathcal{H}_m^\kappa)$ is the class of all sequences $(h_1, \dots, h_d) \langle f_1, \dots, f_d, g_1, \dots, g_c, z^{(1)}, \dots, z^{(d)} \rangle$ for $f_i \in \mathcal{H}_m^\kappa$ with $1 \leq i \leq d$ and $g_j \in \mathcal{H}_\ell^\kappa$ with $1 \leq j \leq c$.

We consider $\mathcal{Z}_{\ell, m}^{c, d, 2k}(\mathcal{H}_\ell^{2k}, \mathcal{H}_m^{2k})$ for some fixed $k \in \mathbb{N}, k \geq 1$. For the parameters $d = 2$ and $c = 1$, this is the hash class used by Dietzfelbinger and Woelfel in [32]. We first analyze the properties of this hash class by stating a definition similar to Definition 2.4 and a lemma similar to Lemma 2.5. We hope that comparing the proofs of Lemma 2.5 and Lemma 5.3 shows the (relative) simplicity of the original analysis.

DEFINITION 5.2. *For $T \subseteq U$, define the random variable d_T , the “deficiency” of $\vec{h} = (h_1, \dots, h_d)$ with respect to T , by $d_T(\vec{h}) = |T| - \max\{k, |g_1(T)|, \dots, |g_c(T)|\}$. (Note: d_T depends only on the g_j -components of (h_1, \dots, h_d) .) Further, define*

- (i) bad_T as the event that $d_T > k$;
- (ii) good_T as $\overline{\text{bad}_T}$, i.e., the event that $d_T \leq k$;
- (iii) crit_T as the event that $d_T = k$.

Hash function sequences (h_1, \dots, h_d) in these events are called “ T -bad”, “ T -good”, and “ T -critical”, resp.

The following lemma is identical to [3, Lemma 1].

LEMMA 5.3. *Assume $d \geq 2$, $c \geq 1$, and $k \geq 1$. For $T \subseteq U$, the following holds:*

- (a) $\Pr(\text{bad}_T \cup \text{crit}_T) \leq (|T|^2/\ell)^{ck}$.
- (b) *Conditioned on good_T (or on crit_T), the hash values $(h_1(x), \dots, h_d(x))$, $x \in T$, are distributed uniformly and independently in $[r]^d$.*

5.2. Application of the Hash Class. The central lemma to bound the impact of using our hash class in contrast to fully random hash functions was Lemma 2.11. One can reprove this lemma in an analogous way for the generalized version of the hash class, using the probability bound from Lemma 5.3(a) to get the following result.

LEMMA 5.4. *Let $c \geq 1$, $k \geq 1$, $S \subseteq U$ with $|S| = n$, and let \mathbf{A} be a graph property. Let $\mathbf{B} \supseteq \mathbf{A}$ be a peelable graph property. Let \mathbf{C} be a graph property such that \mathbf{B} is $C\text{-}2ck\text{-reducible}$. Then*

$$\Pr(\mathbf{B}_S^{\mathbf{A}}) \leq \Pr(\mathbf{B}_S^{\mathbf{B}}) \leq \ell^{-ck} \sum_{t=2k}^n t^{2ck} \cdot \mu_t^{\mathbf{C}}.$$

5.3. Discussion. One can now redo all the calculations from Section 3 and Section 4. We discuss the differences. Looking at Lemma 5.4, we notice the (t^{2ck}) -factor in the sum instead of t^{2c} . Since k is fixed, this factor does not change anything in the calculations that always used $t^{O(1)}$ (see, e.g., the proof of Lemma 3.7). The factor $1/\ell^{ck}$ (instead of $1/\ell^c$) leads to lower values for c if $k \geq 2$. E.g., in cuckoo hashing with a stash, we have to set $c \geq (s+2)/(\delta k)$ instead of $c \geq (s+2)/\delta$. This improves the space usage, since we need less tables filled with random values. However, the higher degree of independence needed for the f - and g -components leads to a higher evaluation time of a single function.

6. Conclusion and Open Questions. We have described a general framework for analyzing hashing-based algorithms and data structures whose analysis depends on properties of the random graph $G(S, \vec{h})$, where \vec{h} comes from a certain class \mathcal{Z} of simple hash functions. This class combined lookups in small random tables with the evaluation of simple 2-universal or 2-independent hash functions.

We developed a framework that allowed us to consider what happens to certain hashing-based algorithms or data structures when fully random hash functions are replaced by hash functions from hash class \mathcal{Z} . If the analysis works using the so-called first-moment method in the fully random case, the framework makes it possible to analyze the situation without exploiting details of the hash function construction. Thus, it requires no knowledge of the hash function and only expertise in random graph theory.

Using this framework we showed that hash functions from class \mathcal{Z} can be used in such diverse applications as cuckoo hashing (with a stash), generalized cuckoo hashing, the simulation of uniform hash functions, the construction of a perfect hash function, and load balancing. Particular choices for the parameters to set up hash functions from \mathcal{Z} provide hash functions that can be evaluated efficiently.

We collect some pointers for future work. Our method is tightly connected to the first moment method. Unfortunately, some properties of random graphs cannot be proven using this method. For example, the classical proof that the connected components of the random graph $G(S, h_1, h_2)$ for $m = (1 + \varepsilon)|S|$, for $\varepsilon > 0$, with fully random hash functions have size $O(\log n)$ uses a Galton-Watson process (see, e.g., [5]). From previous work [25, 26] we know that hash class \mathcal{Z} has some classical properties regarding the balls-into-bins game. In the hypergraph setting this translates to a degree distribution of the vertices close to the fully random case. It would be very interesting to see if such a framework is also possible for other hash function constructions such as [61, 11]. The analysis of generalized cuckoo hashing could succeed (asymptotically) using hash functions from \mathcal{Z} . For this, one has to extend the analysis of the behavior of \mathcal{Z} on small connected hypergraphs to connected hypergraphs with super-logarithmically many edges. Witness trees are another approach to tackle the analysis of generalized cuckoo hashing. We presented initial results in Section 4.3. It is open whether this approach yields good bounds on the space utilization of generalized cuckoo hashing. In light of the new constructions of Thorup [71] and Christiani, Pagh, and Thorup [12], it would be interesting to see whether or not highly-independent hash classes with constant evaluation time are efficient in practice. Moreover, it would be nice to prove that hash class \mathcal{Z} allows running linear probing robustly or to show that it is ε -minwise independent (for suitable ε).

REFERENCES

- [1] YURIY ARBITMAN, MONI NAOR, AND GIL SEGEV, *De-amortized cuckoo hashing: Provable worst-case performance and experimental results*, in Proc. of the 36th International Colloquium on Automata, Languages and Programming (ICALP'09), Springer, 2009, pp. 107–118.
- [2] MARTIN AUMÜLLER, MARTIN DIETZFELBINGER, AND PHILIPP WOELFEL, *Explicit and efficient hash families suffice for cuckoo hashing with a stash*, in Proc. of the 20th Annual European Symposium on Algorithms (ESA'12), Springer, 2012, pp. 108–120.
- [3] MARTIN AUMÜLLER, MARTIN DIETZFELBINGER, AND PHILIPP WOELFEL, *Explicit and efficient hash families suffice for cuckoo hashing with a stash*, *Algorithmica*, 70 (2014), pp. 428–456.
- [4] YOSHI AZAR, ANDREI Z. BRODER, ANNA R. KARLIN, AND ELI UPFAL, *Balanced allocations*, *SIAM J. Comput.*, 29 (1999), pp. 180–200.
- [5] BÉLA BOLLOBÁS, *Random Graphs*, Academic Press, London, 1985.
- [6] FABIANO C. BOTELHO, RASMUS PAGH, AND NIVIO ZIVIANI, *Simple and space-efficient minimal perfect hash functions*, in Proc. of the 10th International Workshop on Algorithms and Data Structures (WADS'07), Springer, 2007, pp. 139–150.
- [7] ———, *Practical perfect hashing in nearly optimal space*, *Inf. Syst.*, 38 (2013), pp. 108–131.
- [8] NEIL J. CALKIN, *Dependent sets of constant weight binary vectors*, *Combinatorics, Probability and Computing*, 6 (1997), pp. 263–271.
- [9] J. LAWRENCE CARTER AND MARK N. WEGMAN, *Universal classes of hash functions (extended abstract)*, in Proc. of the 9th Annual ACM Symposium on Theory of Computing (STOC'77), ACM, 1977, pp. 106–112.
- [10] LARRY CARTER AND MARK N. WEGMAN, *Universal classes of hash functions*, *J. Comput. Syst. Sci.*, 18 (1979), pp. 143–154.
- [11] L. ELISA CELIS, OMER REINGOLD, GIL SEGEV, AND UDI WIEDER, *Balls and bins: Smaller hash families and faster evaluation*, *SIAM J. Comput.*, 42 (2013), pp. 1030–1050.
- [12] TOBIAS CHRISTIANI, RASMUS PAGH, AND MIKKEL THORUP, *From independence to expansion and back again*, in Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing (STOC'15), ACM, 2015, pp. 813–820.
- [13] RICHARD COLE, ALAN M. FRIEZE, BRUCE M. MAGGS, MICHAEL MITZENMACHER, ANDRÉA W. RICA, RAMESH K. SITARAMAN, AND ELI UPFAL, *On balls and bins with deletions*, in Proc. of the 2nd International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM'98), Springer, 1998, pp. 145–158.
- [14] RICHARD COLE, BRUCE M. MAGGS, FRIEDHELM MEYER AUF DER HEIDE, MICHAEL MITZENMACHER, ANDRÉA W. RICA, KLAUS SCHRÖDER, RAMESH K. SITARAMAN, AND BERTHOLD VÖCKING, *Randomized protocols for low congestion circuit routing in multistage interconnection networks*, in Proc. of the 30th Annual ACM Symposium on Theory of Computing (STOC'98), ACM, 1998, pp. 378–388.
- [15] ZBIGNIEW J. CZECH, GEORGE HAVAS, AND BOHDAN S. MAJEWSKI, *Perfect hashing*, *Theor. Comput. Sci.*, 182 (1997), pp. 1–143.
- [16] SØREN DAHLGAARD, MATHIAS BÆK TEJS KNUDSEN, EVA ROTENBERG, AND MIKKEL THORUP, *The power of two choices with simple tabulation*, in Proc. of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'16), 2016. To appear.
- [17] SØREN DAHLGAARD AND MIKKEL THORUP, *Approximately minwise independence with twisted tabulation*, in Proc. of the 14th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT'14), Springer, 2014, pp. 134–145.
- [18] LUC DEVROYE AND EBRAHIM MALALLA, *On the k -orientability of random graphs*, *Discrete Mathematics*, 309 (2009), pp. 1476–1490.
- [19] LUC DEVROYE AND PAT MORIN, *Cuckoo hashing: Further analysis*, *Inf. Process. Lett.*, 86 (2003), pp. 215–219.
- [20] REINHARD DIESTEL, *Graph Theory*, Springer, 2005.
- [21] MARTIN DIETZFELBINGER, *Design strategies for minimal perfect hash functions*, in 4th International Symposium on Stochastic Algorithms: Foundations and Applications (SAGA'07), Springer, 2007, pp. 2–17.
- [22] ———, *On randomness in hash functions (invited talk)*, in 29th International Symposium on Theoretical Aspects of Computer Science (STACS'12), Springer, 2012, pp. 25–28.
- [23] MARTIN DIETZFELBINGER, ANDREAS GOERDT, MICHAEL MITZENMACHER, ANDREA MONTANARI, RASMUS PAGH, AND MICHAEL RINK, *Tight thresholds for cuckoo hashing via XORSAT*, in Proc. of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10), Springer, 2010, pp. 213–225.
- [24] MARTIN DIETZFELBINGER, TORBEN HAGERUP, JYRKI KATAJAINEN, AND MARTTI PENTTONEN, *A reliable randomized algorithm for the closest-pair problem*, *J. Algorithms*, 25 (1997), pp. 19–51.
- [25] MARTIN DIETZFELBINGER AND FRIEDHELM MEYER AUF DER HEIDE, *A new universal class of hash*

- functions and dynamic hashing in real time*, in Proc. of the 17th International Colloquium on Automata, Languages and Programming (ICALP'90), Springer, 1990, pp. 6–19.
- [26] ———, *Dynamic hashing in real time*, in Informatik, Festschrift zum 60. Geburtstag von Günter Hotz, Teubner, 1992, pp. 95–119.
 - [27] ———, *Simple, efficient shared memory simulations*, in Proc. of the 5th ACM Symposium on Parallelism in Algorithms and Architectures, (SPAA'93), ACM, 1993, pp. 110–119.
 - [28] MARTIN DIETZFELBINGER AND MICHAEL RINK, *Applications of a splitting trick*, in Proc. of the 36th International Colloquium on Automata, Languages and Programming (ICALP'09), Springer, 2009, pp. 354–365.
 - [29] MARTIN DIETZFELBINGER AND ULF SCHELLBACH, *On risks of using cuckoo hashing with simple universal hash classes*, in Proc. of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'09), SIAM, 2009, pp. 795–804.
 - [30] ———, *Weaknesses of cuckoo hashing with a simple universal hash class: The case of large universes*, in Proc. of the 35th Conference on Current Trends in Theory and Practice of Informatics (SOFSEM'09), 2009, pp. 217–228.
 - [31] MARTIN DIETZFELBINGER AND CHRISTOPH WEIDLING, *Balanced allocation and dictionaries with tightly packed constant size bins*, Theor. Comput. Sci., 380 (2007), pp. 47–68.
 - [32] MARTIN DIETZFELBINGER AND PHILIPP WOELFEL, *Almost random graphs with simple hash functions*, in Proc. of the 35th Annual ACM Symposium on Theory of Computing (STOC'03), ACM, 2003, pp. 629–638.
 - [33] DAVID EPPSTEIN, MICHAEL T. GOODRICH, MICHAEL MITZENMACHER, AND PAWEŁ PSZONA, *Wear minimization for cuckoo hashing: How not to throw a lot of eggs into one basket*, in Proc. of the 13th International Symposium of Experimental Algorithms, (SEA'14), Springer, 2014, pp. 162–173.
 - [34] PAUL ERDŐS AND ALFRÉD RÉNYI, *On the evolution of random graphs*, Publ. Math. Inst. Hung. Acad. Sci, 5 (1960), pp. 17–61.
 - [35] DIMITRIS FOTAKIS, RASMUS PAGH, PETER SANDERS, AND PAUL G. SPIRAKIS, *Space efficient hash tables with worst case constant access time*, Theory Comput. Syst., 38 (2005), pp. 229–248.
 - [36] NIKOLAOS FOUNTOLAKIS AND KONSTANTINOS PANAGIOTOU, *Orientability of random hypergraphs and the power of multiple choices*, in Proc. of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10), Springer, 2010, pp. 348–359.
 - [37] NIKOLAOS FOUNTOLAKIS, KONSTANTINOS PANAGIOTOU, AND ANGELIKA STEGER, *On the insertion time of cuckoo hashing*, SIAM J. Comput., 42 (2013), pp. 2156–2181.
 - [38] EDWARD A. FOX, LENWOOD S. HEATH, QI FAN CHEN, AND AMJAD M. DAOUD, *Practical minimal perfect hash functions for large databases*, Commun. ACM, 35 (1992), pp. 105–121.
 - [39] MICHAEL L. FREDMAN, JÁNOS KOMLÓS, AND ENDRE SZEMERÉDI, *Storing a sparse table with $o(1)$ worst case access time*, J. ACM, 31 (1984), pp. 538–544.
 - [40] ALAN M. FRIEZE AND PÁLL MELSTED, *Maximum matchings in random bipartite graphs and the space utilization of cuckoo hash tables*, Random Struct. Algorithms, 41 (2012), pp. 334–364.
 - [41] ALAN M. FRIEZE, PÁLL MELSTED, AND MICHAEL MITZENMACHER, *An analysis of random-walk cuckoo hashing*, SIAM J. Comput., 40 (2011), pp. 291–308.
 - [42] PIOTR INDYK, *A small approximately min-wise independent family of hash functions*, J. Algorithms, 38 (2001), pp. 84–90.
 - [43] JIAYANG JIANG, MICHAEL MITZENMACHER, AND JUSTIN THALER, *Parallel peeling algorithms*, in Proc. of the 26th ACM Symposium on Parallelism in Algorithms and Architectures, (SPAA'14), ACM, 2014, pp. 319–330.
 - [44] MICHAŁ KAROŃSKI AND TOMASZ LUCZAK, *The phase transition in a random hypergraph*, Journal of Computational and Applied Mathematics, 142 (2002), pp. 125–135.
 - [45] RICHARD M. KARP, MICHAEL LUBY, AND FRIEDHELM MEYER AUF DER HEIDE, *Efficient PRAM simulation on a distributed memory machine*, Algorithmica, 16 (1996), pp. 517–542.
 - [46] ———, *Efficient PRAM simulation on a distributed memory machine*, Algorithmica, 16 (1996), pp. 517–542.
 - [47] MEGHA KHOSLA, *Balls into bins made faster*, in Proc. of the 21st Annual European Symposium on Algorithms (ESA'13), Springer, 2013, pp. 601–612.
 - [48] ADAM KIRSCH, MICHAEL MITZENMACHER, AND UDI WIEDER, *More robust hashing: Cuckoo hashing with a stash*, in Proc. of the 16th Annual European Symposium on Algorithms (ESA'08), Springer, 2008, pp. 611–622.
 - [49] ———, *More robust hashing: Cuckoo hashing with a stash*, SIAM J. Comput., 39 (2009), pp. 1543–1561.
 - [50] TORYN QWYLLYN KLASSEN AND PHILIPP WOELFEL, *Independence of tabulation-based hash classes*, in Proc. Theoretical Informatics - 10th Latin American Symposium (LATIN'12), Springer, 2012, pp. 506–517.

- [51] BOHDAN S. MAJEWSKI, NICHOLAS C. WORMALD, GEORGE HAVAS, AND ZBIGNIEW J. CZECH, *A family of perfect hashing methods*, Comput. J., 39 (1996), pp. 547–554.
- [52] FRIEDHELM MEYER AUF DER HEIDE, CHRISTIAN SCHEIDELER, AND VOLKER STEMANN, *Exploiting storage redundancy to speed up randomized shared memory simulations*, Theor. Comput. Sci., 162 (1996), pp. 245–281.
- [53] MARC MEZARD AND ANDREA MONTANARI, *Information, Physics, and Computation*, Oxford University Press, 2009.
- [54] MICHAEL MITZENMACHER AND SALIL P. VADHAN, *Why simple hash functions work: exploiting the entropy in a data stream*, in Proc. of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA’08), SIAM, 2008, pp. 746–755.
- [55] MICHAEL MOLLOY, *Cores in random hypergraphs and boolean formulas*, Random Struct. Algorithms, 27 (2005), pp. 124–135.
- [56] ANNA ÖSTLIN AND RASMUS PAGH, *Uniform hashing in constant time and linear space*, in Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC’03), ACM, 2003, pp. 622–628.
- [57] RICHARD OTTER, *The number of trees*, Annals of Mathematics, (1948), pp. 583–599.
- [58] ANNA PAGH AND RASMUS PAGH, *Uniform hashing in constant time and optimal space*, SIAM J. Comput., 38 (2008), pp. 85–96.
- [59] ANNA PAGH, RASMUS PAGH, AND MILAN RUZIC, *Linear probing with constant independence*, SIAM J. Comput., 39 (2009), pp. 1107–1120.
- [60] RASMUS PAGH AND FLEMMING FRICHE RODLER, *Cuckoo hashing*, J. Algorithms, 51 (2004), pp. 122–144.
- [61] MIHAI PĂTRAȘCU AND MIKKEL THORUP, *The power of simple tabulation hashing*, J. ACM, 59 (2012), p. 14.
- [62] ———, *Twisted tabulation hashing.*, in Proc. of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA’13), SIAM, 2013, pp. 209–228.
- [63] ———, *On the k -independence required by linear probing and minwise independence*, ACM Trans. Algorithms, 12 (2015), pp. 8:1–8:27.
- [64] SANGUTHEVAR RAJASEKARAN, PANOS M. PARDALOS, JOHN H. REIF, AND ROSÉ ROLIM, eds., *Handbook of Randomized Computing, Vol. 1*, Kluwer Academic Publishers, 2001.
- [65] OMER REINGOLD, RON D. ROTHBLUM, AND UDI WIEDER, *Pseudorandom graphs in data structures*, in Proc. of the 41st International Colloquium on Automata, Languages and Programming (ICALP’14), Springer, 2014, pp. 943–954.
- [66] MICHAEL RINK, *Thresholds for Matchings in Random Bipartite Graphs with Applications to Hashing-Based Data Structures*, PhD thesis, Technische Universität Ilmenau, 2014.
- [67] THOMAS SCHICKINGER AND ANGELIKA STEGER, *Simplified witness tree arguments*, in Proc. of the 27th Conference on Current Trends in Theory and Practice of Informatics (SOFSEM’00), Springer, 2000, pp. 71–87.
- [68] JEANETTE SCHMIDT-PRUZAN AND ELI SHAMIR, *Component structure in the evolution of random hypergraphs*, Combinatorica, 5 (1985), pp. 81–94.
- [69] ALAN SIEGEL, *On universal classes of extremely random constant-time hash functions*, SIAM J. Comput., 33 (2004), pp. 505–543.
- [70] VOLKER STEMANN, *Parallel balanced allocations*, in Proc. of the 8th ACM Symposium on Parallelism in Algorithms and Architectures, (SPAA’96), ACM, 1996, pp. 261–269.
- [71] MIKKEL THORUP, *Simple tabulation, fast expanders, double tabulation, and high independence*, in Proc. 54th Annual Symposium on Foundations of Computer Science (FOCS), ACM, 2013, pp. 90–99.
- [72] MIKKEL THORUP AND YIN ZHANG, *Tabulation-based 5-independent hashing with applications to linear probing and second moment estimation*, SIAM J. Comput., 41 (2012), pp. 293–331.
- [73] BERTHOLD VÖCKING, *How asymmetry helps load balancing*, J. ACM, 50 (2003), pp. 568–589.
- [74] MARK N. WEGMAN AND LARRY CARTER, *New classes and applications of hash functions*, in Proc. 20th Annual Symposium on Foundations of Computer Science (FOCS’79), IEEE Computer Society, 1979, pp. 175–182.
- [75] ———, *New hash functions and their use in authentication and set equality*, J. Comput. Syst. Sci., 22 (1981), pp. 265–279.
- [76] PHILIPP WOELFEL, *Efficient strongly universal and optimally universal hashing*, in 24th International Symposium on Mathematical Foundations of Computer Science (MFCS’99), Springer, 1999, pp. 262–272.
- [77] ———, *Asymmetric balanced allocation with simple hash functions*, in Proc. of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA’06), ACM, 2006, pp. 424–433.

Appendix A. Proof of Lemma 4.22.

Proof. For the sake of the analysis, we regard $\text{MCWT}(k, t)$ as the union of two graph properties $\text{MC}(k, t)$, hypergraphs that form (k, t) -multicycles, and $\text{WT}(t-1)$, hypergraphs that form witness trees for the parameter $t-1$. We show the lemma by proving $E^*(N_S^{\text{MC}(k, t)}) = O(n^{-\alpha})$ and $E^*(N_S^{\text{WT}(t-1)}) = O(n^{-\alpha})$. In both cases, we consider the bipartite representation of hypergraphs. Our proofs follow [67, Section 4].

We start by bounding $E^*(N_S^{\text{MC}(k, t)})$. As we have seen, a (k, t) -multicycle is a connected graph that has at most $2kt$ vertices and cyclomatic number k . We start by counting (k, t) -multicycles with exactly s vertices and $s + k - 1$ edges, for $s \leq 2kt$. In this case, we have to choose j and u (the number of jobs and machines, resp.) such that $s = j + u$. By Lemma 3.2 there are at most $(s + k - 1)^{O(k)}$ unlabeled (k, t) -multicycles. Fix such an unlabeled (k, t) -multicycle G . There are two ways to split the vertices of G into the two sides of the bipartition. (G can be assumed to be bipartite since we consider (k, t) -multicycles that are subgraphs of the allocation graph.) Once this bipartition is fixed, we have n^j ways to choose the job vertices and label vertices of G with these jobs. There are d^{s+k-1} ways to label the edges of G with labels from $1, \dots, d$, which represent the request modeled by an edge between a job vertex and a machine vertex. Once this labeling is fixed, there are $(n/d)^u$ ways to choose machine vertices and label the remaining vertices of G . Fix such a fully labeled graph G' .

For each request r of a job $w = 1, \dots, j$, choose a machine from $[n/d]$ at random and independently. The probability that this machine is the same machine that w had chosen in G' is d/n . Thus, the probability that G' is realized by the random choices is $(d/n)^{s+k-1}$. By setting $k = \alpha + 2$ and using the parameter choice $t = O(\ln \ln n)$ we calculate

$$\begin{aligned} E^*(N_S^{\text{MC}(k, t)}) &\leq \sum_{s=1}^{2kt} \sum_{u+j=s} 2 \cdot n^j \cdot (n/d)^u \cdot d^{s+k-1} \cdot (s+k-1)^{O(k)} \cdot (d/n)^{s+k-1} \\ &\leq n^{1-k} \sum_{s=1}^{2kt} 2s \cdot d^{2(s+k-1)} \cdot (s+k-1)^{O(1)} \\ &\leq n^{1-k} \cdot 2kt \cdot 4kt \cdot d^{2(2kt+k-1)} \cdot (2kt+k-1)^{O(1)} \\ &\leq n^{1-k} \cdot (\ln \ln n)^{O(1)} \cdot (\ln n)^{O(1)} = O(n^{2-k}) = O(n^{-\alpha}). \end{aligned}$$

Now we consider $E^*(N_S^{\text{WT}(t-1)})$. By the simple recursive structure of witness trees, a witness tree of depth $t-1$ has $j = \frac{\tau^{t-1}(d-1)^{t-2}-\tau}{\tau(d-1)-1}$ job vertices and $u = \frac{\tau^{t-1}(d-1)^{t-1}-1}{\tau(d-1)-1}$ machine vertices. Let T be an unlabeled witness tree of depth $t-1$. T has $r = d \cdot j$ edges. There are at most n^j ways to choose j jobs from S and label the job vertices of T and at most d^r ways to label the edges with a label from $\{1, \dots, d\}$. Once this labeling is fixed, there are at most $(n/d)^u$ ways to choose the machines and label the machine vertices in the witness tree. With these rough estimates, we over-counted the number of witness trees by at least a factor of $(\tau!)^{j/\tau} \cdot ((d-1)!)^j$. (See Figure 4.1. For each job vertex, there are $(d-1)!$ labelings which result in the same witness tree. Furthermore, for each non-leaf machine vertex, there are $\tau!$ many labelings which yield the same witness tree.) Fix such a fully labeled witness tree T' .

For each request of a job $w = 1, \dots, j$ choose at random a machine from $[n/d]$. The probability that the edge matches the edge in T' is d/n . Thus, the probability

that T' is realized by the random choices is $(d/n)^r$. We calculate

$$\begin{aligned}
\mathbb{E}^*(N_S^{\text{WT}(t-1)}) &\leq n^j \cdot d^r \cdot (n/d)^u \cdot \left(\frac{1}{\tau!}\right)^{j/\tau} \cdot \left(\frac{1}{(d-1)!}\right)^j \cdot (d/n)^r \\
&\leq n \cdot d^{2r} \cdot \left(\frac{1}{\tau!}\right)^{j/\tau} \cdot \left(\frac{1}{(d-1)!}\right)^j \leq n \left[\frac{e}{\tau} \cdot \left(\frac{e}{d-1}\right)^{d-1} \cdot d^{2d} \right]^j \\
&\leq n \left(\frac{e^d \cdot d^{d+1}}{\tau} \right)^j.
\end{aligned}$$

Observe that

$$j = \frac{\tau^{t-1}(d-1)^{t-2} - \tau}{\tau(d-1) - 1} \geq \frac{\tau^{t-2}(d-1)^{t-2} - \tau}{d} \geq \frac{(\tau(d-1))^{t-2}}{2d}.$$

For the parameter settings assumed in the lemma we get

$$\begin{aligned}
\mathbb{E}^*(N_S^{\text{WT}(t-1)}) &\leq n \left(\frac{e^d \cdot d^{d+1}}{\tau} \right)^{\frac{(\tau(d-1))^{t-2}}{2d}} = n \left(\frac{e^d \cdot d^{d+1}}{\tau} \right)^{\frac{\beta t \ln n}{2d \ln \ln n}} \\
&\leq n \left(e^d \cdot d^{d+2} \cdot \left(\frac{\beta \ln \ln n}{t \ln n} \right)^{\frac{1}{t-2}} \right)^{\frac{\beta t \ln n}{2d \ln \ln n}} \\
&\leq n \cdot \left((e^d \cdot d^{d+2})^t \left(\frac{\beta \ln \ln n}{t \ln n} \right) \right)^{\frac{\beta \ln n}{2d \ln \ln n}} \\
&\leq n \cdot \left((e^d \cdot d^{d+2})^{\frac{1}{\beta} \ln \ln n} \cdot \frac{1}{\ln n} \right)^{\frac{\beta \ln n}{2d \ln \ln n}} \\
&\leq n^{3/2 + \ln d - \beta/2d}.
\end{aligned}$$

Setting $\beta = 2d(\alpha + \ln d + 3/2)$ suffices to show that $\mathbb{E}^*(N_S^{\text{MCWT}(k,t)}) = O(n^{-\alpha})$. \square